"In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institution shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

"

5a

12R

ELEMENTS OF LINEAR REGRESSION: AN EXPOSITORY DEVELOPMENT

A THESIS

Presented to

the Faculty of the Graduate Division

by

Ollie Brown Francis, Jr.

In Partial Fulfillment of the

Requirements for the Degree

Master of Science in Applied Mathematics

Georgia Institute of Technology

June 1960

ELEMENTS OF LINEAR REGRESSION: AN

EXPOSITORY DEVELOPMENT

Approved:

J. W. Walker, Chairman

R. H. Kasriel

J. J. Moder

Date Approved by Chairman: 7 June 60

ACKNOWLEDGEMENT

The author wishes to express his sincere appreciation to his thesis advisor, Dr. James W. Walker, both for suggesting this topic and for his guidance during the course of this study.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENT	ii
LIST OF TABLES	iv
Chapter	
I. INTRODUCTION	1
II. LEAST SQUARES	13
III. STATISTICAL ESTIMATION OF THE PARAMETERS IN THE REGRESSION MODEL	24
IV. TESTING HYPOTHESES AND ANALYSIS OF VARIANCE	34
V. COMPUTATIONAL PROCEDURES	51
APPENDIX. THE NORMAL AND RELATED DISTRIBUTIONS	77
BIBLIOGRAPHY	84

LIST OF TABLES

Table | Page

1. Analysis of Variance Table in Terms of Orthonormal Variables 44
2. Condensed Analysis of Variance Table in Terms of Orthonormal Variables 45
3. The Gaussian Tableaux 57
4. Percentage of Certain Materials Found in Tobacco Leaf 70
5. The Gaussian Tableaux for the Tobacco Leaf Problem 71

CHAPTER I

INTRODUCTION

The aim of this paper is to present a unified development of the elements of regression theory. This paper deals with linear regression only. For a discussion of non-linear regression reference may be made to Williams (8), Chapter 4.

In this chapter a definition is given for the linear regression model, and an attempt is made to provide a connection between this model and multivariate distributions. In the second chapter the appropriate parts of least squares theory are developed in a manner applicable to the subsequent statistical estimation and testing of parameters in the regression model. In Chapter III we show that certain "best" statistical estimates of the parameters in the regression model are formally identical with the least squares estimates obtained in Chapter II. In Chapter IV, invoking the assumption of normality of distribution of errors, the fundamental tests on the regression parameters are formally developed. Chapter V contains the algorithms necessary for obtaining both the estimates and tests of hypotheses described in the preceding two chapters.

Definition 1. We shall say that we have a regression model in case there are N random variables $X_1, \ldots, x_N$ which are representable as linear combinations of p + 1 unknown quantities $\beta_o, \beta_1, \ldots \beta_p$, plus random errors $\varepsilon_1, \ldots, \varepsilon_N$

$$x_\mu = \beta_o + \sum_{i=1}^{p} z_{\mu i}\, \beta_i \quad \varepsilon_\mu \quad , \mu = 1, 2, \ldots, N \tag{1.1}$$

where the $z_{\mu i}$ are known constants. This is essentially the definition given by Scheffé (6), page 4, for the mathematical model used in both analysis of variance and regression analysis.

We shall denote this model in the equivalent form

$$x = \beta_o z_o + \beta_1 z_1 + \cdots + \beta_p z_p + \varepsilon \tag{1.1a}$$

where

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad z_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad z_i = \begin{bmatrix} z_{1i} \\ \vdots \\ z_{Ni} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{1.1b}$$

or in the form $x = z\beta + \varepsilon$

$$z = [z_o \cdots z_p], \quad \beta = \begin{bmatrix} \beta_o \\ \vdots \\ \beta_p \end{bmatrix}$$

We shall also occasionally write (1.1a) to represent the random variable x generically as a function of the variables $z_o, z_1, \ldots, z_p$ and ε.

We observe that, while x and ε are random vectors, the so-called structural vectors $z_0, \ldots, z_p$ are nonrandom.

Certain minimal assumptions are usually made regarding this model. Once and for all we make the following assumptions:

(a) $E\varepsilon_\mu = 0$ for $\mu = 1, \ldots, N$

(b) The structural vectors are linearly independent.[1] (This automatically requires $N > p$.)

We shall in the sequel add one or two additional restrictions.

---

[1]This assumption is non-restrictive. If the structural vectors are linearly dependent, the problem can be reduced to one where they are linearly independent. This conversion, however, is nonunique. See Scheffé (6), Chapter 1, for a discussion of this problem.

The range of applicability of the regression model as defined above covers such matters as testing the equality of means of several approximately normal populations or testing the significance of factors in various experimental designs (in which the structural vectors, or controllable vectors, indicate frequently merely the presence or absence of a factor at some "level"). We shall not discuss the nature and classification of the areas of application here. Reference can be made to Scheffé (6), Chapter 1, Fraser (3), Chapter 9, and other standard works.

Although it is not always the case, it is frequently convenient, or consistent with the practical circumstances, to conceive of the random variable x as being determined to within ε by the condition of knowing the values $z_1, z_2, \ldots, z_p$ of p other random variables. From this point of view x is a variable whose conditional mean is some function of the z_i.

Following Cramér (2) we define below the notions of a regression surface and a mean square regression plane.

Let $x, z_1, \ldots, z_p$ be $p + 1$ random variables with continuous joint probability density function $f(x, z_1, \ldots, z_p)$. Consider the conditional density function

$$f(x|z_1,\ldots,z_p) = \frac{f(x, z_1,\ldots,z_p)}{\int_{-\infty}^{\infty} f(x',z_1,\ldots,z_p)\, dx'}$$

and the conditional mean value of x

$$E(x|z_1\ldots z_p) = \int_{-\infty}^{\infty} xf(x|z_1,\ldots,z_p)\, dx \tag{1.2}$$

$$= \int_{-\infty}^{\infty} x \frac{f(x,z_1,\ldots,z_p)\, dx}{\int_{-\infty}^{\infty} f(x', z_1,\ldots,z_p)\, dx'} = m(z_1,\ldots,z_p)$$

The locus of points $(x^*, z_1, \ldots, z_p)$ in Euclidean $(p + 1)$-space, given by the relation

$$x^* = m(z_1, \ldots, z_p)$$

is defined as the regression surface for the mean of the random variable x. If x^* is linear in the real variables z_i, then the equations in the regression model (1.1) correspond to representations of observations x_μ, on the random variable x, as the conditional mean

$$m_\mu = \beta_o + \beta_1 z_{\mu 1} + \ldots + \beta_p z_{\mu p}$$

plus an error term ε_μ whose expected value must be zero.

Again, let $x, z_1, \ldots, z_p$ have a distribution with finite second order moments. A mean square regression plane for x with respect to $z_1, \ldots, z_p$ is that set of points $(\hat{x}, z_1, \ldots z_p)$ in $(p + 1)$-space satisfying the linear condition

$$\hat{x} = \beta_o + \beta_1 z_1 + \ldots + \beta_p z_p$$

where $\beta_o \ldots \beta_p$ are values of $\alpha_o \ldots \alpha_p$ such that the integral

$$E(x - \alpha_o - \alpha_1 z_1 - \ldots - \alpha_p z_p)^2$$

is a minimum. In such a case $\hat{x}$ is said to be a best linear estimate of x in the sense that, of all linear functions $\alpha_o + \alpha_1 z_1 + \ldots + \alpha_p z_p$ of $z_1, \ldots, z_p$, it is one for which the expected value of the square of the difference between x and the linear estimate is least.

If we suppose $x_1, \ldots, x_N$ are random variables whose expected deviation from the mean square regression plane is zero, the regression model again is appropriate, although the interpretation is not quite the same as the interpretation just given above; for, in general, the conditional expectation of x need not be equal to the value on the mean square regression plane.

Next, we show that under very general conditions the mean square regression plane exists uniquely, and we establish the fact that, if the regression surface is linear, it coincides with the mean square regression plane.

Definition 2. The random variables $x_1, \ldots, x_n$ are said to have a nonsingular distribution in case their covariance matrix (σ_{ij}) where

$$\sigma_{ij} = E(x_i - Ex_i)(x_j - Ex_j)$$

is nonsingular.

Lemma: If $\sigma = (\sigma_{ij})$ is the covariance matrix of a nonsingular distribution, then σ is positive definite.

Proof: Let $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$ be an arbitrary vector. Then

$$\alpha^T \sigma \alpha = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \sigma_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j E(x_i - Ex_i)(x_j - Ex_j)$$

$$= \int \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j (x_i - Ex_i)(x_j - Ex_j)\, dP = \int \left[\sum_{i=1}^{n} \alpha_i (x_i - Ex_i) \right]^2 dP \geq 0$$

Since $\left[\sum_{i=1}^{n} \alpha_i (x_i - Ex_i) \right]^2$ is nonnegative, equality holds if and only if

$$\sum_{i=1}^{n} \alpha_i (x_i - Ex_i) = 0$$

except on a set of probability measure zero. Then for $j = 1, \ldots, n$,

$$\sum_{i=1}^{n} \alpha_i (x_i - Ex_i)(x_j - Ex_j) = 0$$

except on a set of probability measure zero. Then

$$\sum_{i=1}^{n} \alpha_i \int (x_i - Ex_i)(x_j - Ex_j)\, dP = 0$$

Therefore, $\alpha^T \sigma = 0$. Since σ is nonsingular

$$\alpha^T = \alpha^T \sigma \sigma^{-1} = 0\, \sigma^{-1} = 0 \ .$$

Thus

$$\alpha^T \sigma \alpha = 0$$

for every α , with equality holding if and only if

$$\alpha = 0$$

Therefore σ is positive definite.

Theorem 1. If $x, z_1, \ldots, z_p$ have a nonsingular distribution, then a mean square regression plane for x with respect to $z_1, \ldots, z_p$ exists uniquely.

Proof: Write $g(\alpha_0, \ldots, \alpha_p) = E(x - \alpha_0 - \sum_{i=1}^{p} \alpha_i z_i)^2$

Since $g \geq 0$, clearly it is bounded below and hence has a greatest lower bound.

Clearly g is continuous in the vector $\alpha = (\alpha_0, \ldots, \alpha_p)$. We shall now show that

$$\lim_{||\alpha|| \to \infty} g(\alpha) = \infty$$

where $||\alpha||^2 = \alpha^T \alpha$, a norm. In fact

$$g(\alpha) = E\left\{x - Ex + \sum_{i=1}^{p} (-\alpha_i)(z_i - Ez_i) + Ex - \alpha_0 - \sum_{i=1}^{p} \alpha_i Ez_i\right\}^2 =$$

$$= E\left\{\sum_{i=0}^{p} a_i(u_i - Eu_i) + E\, a_i \sum_{i=0}^{p} u_i\right\}^2$$

where $a = \begin{bmatrix} 1 \\ -\alpha_1 \\ \vdots \\ -\alpha_p \end{bmatrix}$, $u_o = x - \alpha_o$, $u_i = z_i$ $(i = 1,\ldots,p)$. So

$$g(\alpha) = E\left[(u - Eu) + Eu\right]^2$$

where $u = \sum_{i=0}^{p} a_i u_i$, and

$$g(\alpha) = E(u - Eu)^2 + (Eu)^2 = \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \sigma_{ij} + (Eu)^2$$

where $\sigma_{ij} = E(U_i - Eu_i)(u_j - Eu_j)$. Therefore

$$g(\alpha) = a^T (\sigma_{ij})\, a + (Eu)^2$$

By hypothesis (σ_{ij}) is nonsingular. Hence, applying the lemma, (σ_{ij}) is positive definite. Applying an orthogonal transformation,

$$g(\alpha) = \gamma^T d\gamma + (Eu)^2$$

where d is diagonal with characteristic roots (positive) in the diagonal. Also, $||\gamma|| = ||a||$. Now as $||\alpha|| \to \infty$, either $||a|| \to \infty$, or $|\alpha_o| \to \infty$ while $\alpha_1,\ldots,\alpha_p$ remain bounded. In the first case $g(\alpha) \to \infty$, since

$$g(\alpha) = \gamma^T d\gamma + (Eu)^2 \geq d_{min}\, \gamma^T \gamma = d_{min}\, ||a||^2$$

In the second case $Eu = Ex - \alpha_o - \sum_{i=1}^{p} \alpha_i Ez_i$ becomes unbounded since α_o becomes unbounded and $Ex - \sum_{i=1}^{p} \alpha_i Ez_i$ remains bounded. Hence $(Eu)^2 \to \infty$, and therefore in this case $g \to \infty$.

Let $B = \underset{\alpha}{\text{glb}}\, g(\alpha)$. It is obvious that $B < \infty$. Let $S = \{\alpha : g(\alpha) \leq B+\varepsilon\}$ where $0 < \varepsilon < \infty$. The set S is bounded, for if it were not, we could take a sequence of points $\{\alpha^{(n)}\}$ in S such that

$$\lim_{n\to\infty} ||\alpha(n)|| = +\infty$$

But this implies that $\lim_{n\to\infty} g(\alpha^{(n)}) = +\infty$, which contradicts the fact that $g(\alpha) \leq B+\varepsilon < \infty$. Thus we can enclose S in a closed interval $I = \{\alpha : a_i \leq \alpha_i \leq b_i\}$.

Since g is a continuous function on a bounded closed set I, g takes on its greatest lower bound at some point in I; and that greatest lower bound is obviously B, since I contains S, and S contains points for which g is arbitrarily close to B. Hence g reaches an absolute minimum at some point of continuity.

It follows, since g is differentiable everywhere, that its first partial derivatives must vanish. It will be shown that these partial derivatives vanish for only one value of α and hence the point where the derivatives vanish is the point at which g is a minimum. Call this point $\beta = (\beta_0 \ldots \beta_p)$. Thus

$$\frac{\partial g}{\partial \alpha_0} = -2E(x - \beta_0 - \sum_{i=1}^{p} \beta_i z_i) = 0$$

$$\frac{\partial g}{\partial \alpha_i} = -2Ez_i(x - \beta_o - \sum_{j=1}^{p} \beta_j z_j) = 0, \; i = 1,\ldots,p$$

or

$$\beta_o = Ex - \sum_{i=1}^{p} \beta_i Ez_i$$

where β_i is such that

$$E\left\{z_i(x - Ex + \sum_{j=1}^{p} \beta_j Ez_j - \sum_{j=1}^{p} \beta_j z_j)\right\} = 0, \ i = 1,\ldots,p$$

or

$$E\left\{(z_i - Ez_i)(x - Ex + \sum_{j=1}^{p} \beta_j Ez_j - \sum_{j=1}^{p} \beta_j z_j)\right\} = 0$$

since

$$E\left\{(Ez_i)(x - Ex + \sum_{j=1}^{p} \beta_j Ez_j - \sum_{j=1}^{p} \beta_j z_j)\right\} =$$

$$= (Ez_i)(Ex - Ex + \sum_{j=1}^{p} \beta_j Ez_j - \sum_{j=1}^{p} \beta_j Ez_j) = 0$$

That is, β is such that

$$\sum_{j=1}^{p} \beta_j \sigma_{ij} = \sigma_{oi}, \ i = 1,\ldots,p$$

$$\beta_o = m_o - \sum_{j=1}^{p} \beta_j m_j$$

where

$$m_o = Ex$$

$$m_i = Ez_i, \ i = 1,\ldots, p$$

$$\sigma_{oi} = E(x - Ex)(z_i - Ez_i), \ i = 1,\ldots,p$$

$$\sigma_{ij} = E(z_i - Ez_i)(z_j - Ez_j), \ i = 1,\ldots,p, j = 1,\ldots,p$$

In matrix notation,

$$\sigma_o = \sigma\beta$$

where

$$\sigma_o = \begin{bmatrix} 01 \\ \\ 0p \end{bmatrix}, \quad \sigma = \begin{bmatrix} \sigma_{1p} & \cdots & \sigma_{1p} \\ \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

Let $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_p \end{bmatrix}$ be an arbitrary vector, and let $\Gamma = \begin{bmatrix} 0 \\ \gamma_1 \\ \gamma_p \end{bmatrix}$. Then

$$\Gamma^T \sigma\gamma = (\gamma_1 \ldots \gamma_p) \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix} = (0\gamma_1 \ldots \gamma_p) \begin{bmatrix} \sigma_{00} & \sigma_{01} & \cdots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & & \sigma_{1p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p0} & \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} 0 \\ \gamma_1 \\ \vdots \\ \gamma_p \end{bmatrix}$$

$$= \Gamma^T \sum \Gamma \geq 0.$$

Since $\sum$ is positive definite, equality holds if and only if $\Gamma = 0$. However, $\Gamma = 0$ if and only if $\gamma = 0$. Thus $\gamma^T \sigma \gamma \geq 0$, with $\gamma^T \sigma \gamma = 0$ if and only if $\gamma = 0$. Therefore σ is positive definite.

This implies that $\sigma_o = \sigma\beta$ has a unique solution. We have proved that if $x, z_1, \ldots, z_p$ have a nonsingular distribution, there exists a unique mean square regression plane for x with respect to $z_1, \ldots, z_p$; and moreover we have found the equations

$$\sigma\beta = \sigma_o$$

$$\beta_o = m_o - \sum_{i=1}^{p} {}_i m_i$$

for the β_o, $\beta_1, \ldots, \beta_p$ which explicitly describe the mean square regression plane.

Definition 2. The plane of closest fit to a surface $y = f(x_1, \ldots, x_n)$ is the plane $\eta = \alpha_o + \alpha_1 x_1 + \ldots + \alpha_n x_n$ for which $E[f(x_1, \ldots, x_n) - \eta]^2$ is a minimum.

<u>Theorem 2</u>. The mean square regression plane, $\hat{x} = \beta_o + \beta_1 z_1 + \dots + \beta_p z_p$, is the plane of closest fit to the regression surface, $x^* = m(z_1, \dots z_p)$, when the latter exists.

Proof: Suppose $m(z_1, \dots, z_p)$ exists. Let

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad z = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}, \quad m(z) = m(z_1, \dots, z_p)$$

Then

$$E(x - \beta_o - \beta^T z)^2 = E\left[x - m(z) + m(z) - \beta_o - \beta^T z\right]^2 =$$

$$= E\left[x - m(z)\right]^2 + 2E\left[x - m(z)\right]\left[m(z) - \beta_o - \beta^T z\right] + E\left[(m(z) - \beta_o - \beta^T z\right]^2$$

However, from (1.2),

$$E\left[x - m(z)\right]\left[m(z) - \beta_o \quad \beta^T z\right]$$

$$= \int_{-\infty}^{\infty}\left[x - m(z)\right]\left[m(z) - \beta_o - \beta^T z\right] f(x,z)\, dx\, dz$$

$$= \int_{-\infty}^{\infty} f_2(z)\left[m(z) - \beta_o - \beta^T z\right]\left[\int_{-\infty}^{\infty}\left[x - m(z)\right] f(x|z)\, dx\right] dz$$

$$= \int_{-\infty}^{\infty} f_2(z)\left[m(z) - \beta_o - \beta^T z\right]\left[\int_{-\infty}^{\infty} x f(x|z)\, dx - m(z)\int_{-\infty}^{\infty} f(x|z)\, dx\right] dz$$

$$= \int_{-\infty}^{\infty} f_2(z)\left[m(z) - \beta_o - \beta^T z\right]\left[m(z) - m(z)\right] dz = 0$$

Here we have written $f(x,z) = f(x,z)\, f_2(z)$, $f_2(z)$ being the marginal probability density function of z.

Therefore

$$E(x - \beta_o - \beta^T z)^2 = E(x - m(z))^2 + E(m(z) - \beta_o - \beta^T z)^2$$

Since $E(x - m(z))^2$ is independent of $\beta_o, \ldots, \beta_p$, $E(x - \beta_o - \beta^T z)^2$ and $E(m(z) - \beta_o - \beta^T z)^2$ are minimized by the same choice of $\beta_o, \ldots, \beta_p$. Thus the mean square regression plane is the plane of closest fit to the regression surface for the mean of x.

<u>Corollary</u>. If the regression surface is a plane, it is the mean square regression plane.

Proof: Suppose $m(z) = \beta_o + \beta^T z$, where $\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$

Then $E(\beta_o + \beta^T z - \alpha_o - \alpha^T z)^2$ is minimized by setting $\alpha_i = \beta_i (i = 0, \ldots, p)$ to get an expected value of zero.

CHAPTER II

LEAST SQUARES

The statistical problems in regression theory arise in connection with the estimation of and testing hypotheses concerning the parameters β_o, $\beta_1,\ldots,\beta_p$ in the regression model as it was defined in Chapter I. Of course the heart of the estimation problem is to obtain estimates with more or less optimum properties. We withhold further discussion of such desirable properties until Chapter III.

In the present chapter the vector matrix

$$\left[x,\ z_1,\ldots,\ z_p\right]$$

of the regression model (1.1a) is viewed as a set of N points or observations $(x_\mu,\ z_{\mu 1},\ldots,\ z_{\mu p})$. The problem posed is to find, among all linear functions,

$$\beta_o + \beta_1\ \tilde{z}_1 + \cdots + \beta_p\ \tilde{z}_p$$

that one, $\hat{x} = b_o + b_1\ \tilde{z}_1 + \cdots + b_p\ \tilde{z}_p$, which minimizes the sum of squares of the differences between x_μ and $\beta_o + \beta_1\ z_{\mu 1} + \cdots + \beta_p\ z_{\mu p}$, for $\mu = 1,2,\ldots,N$. The fundamental theorem of this chapter states that, if z_o, $z_1,\ldots,z_p$ are linearly independent (as assumed), then $S\,\varepsilon_\mu^{\,2} = \varepsilon^T\varepsilon$ is uniquely minimized by

$$B = H^{-1}\ G \tag{2.1}$$

where $H = z^T z$ and $B = z^T x$.

It is interesting to note that this theorem is really a corollary to Theorem 1 in Chapter I. The proof follows immediately if the joint distribution of random variables $x, z_1, \ldots, z_p$ is defined so that a mass of $1/N$ is placed at each of the points $(x_\mu, z_{\mu 1}, \ldots, z_{\mu p})$. One merely has to carry out the details of taking expected values to obtain (2.1).

We prefer, however, to give a different proof - one which involves the orthogonal transformations which are important to the subsequent subject of statistical tests. The proof requires no calculus theory.

Definition 1. We shall call

$$x = z' \beta' + \varepsilon$$

an equivalent form of the model (1.1) in case

$$z' = z\alpha, \quad \beta' = \alpha^{-1}\beta$$

For a fixed set of β_i it is clear that the equivalent forms of the model (1.1) all have the same $\varepsilon^T\varepsilon$.

Theorem 1. Corresponding to any model

$$x = z\beta + \varepsilon,$$

where $\{z_i\}$ are linearly independent, is a completely orthogonal equivalent form

$$x = z^* \beta^* + \varepsilon$$

where $(z_i^* \cdot z_j^*) = 0$ for $i \neq j$

and $z^* = z\alpha$, $\alpha = \begin{bmatrix} 1 & \alpha_{01} & \alpha_{02} \cdots & \alpha_{0p} \\ 0 & 1 & \alpha_{12} \cdots & \alpha_{1p} \\ 0 & 0 & 1 \quad \cdots & \alpha_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \quad \cdots & 1 \end{bmatrix}$

Proof: We make a preliminary orthogonalization:

Let $z' = z\alpha'$ where $\alpha' = \begin{bmatrix} 1 & -\bar{z}_1 & -\bar{z}_2 \dots & -\bar{z}_p \\ 0 & 1 & 0 \dots & 0 \\ 0 & 0 & 1 \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & 1 \end{bmatrix}$

Then $z'_{\mu o} = z_{\mu o} = 1$

$$z'_{\mu i} = z_{\mu i} - \bar{z}_i, \ i = 1,\dots,p$$

$$(z'_o \cdot z'_i) = Sz'_{\mu o}\, z'_{\mu i} = Sz'_{\mu i} = S(z_{\mu i} - \bar{z}_i) = Sz_{\mu i} - N\bar{z}_i = 0$$

Thus z'_o is orthogonal for each z'_i, $i = 1,\dots,p$.

For the vectors $z'_1 \dots z'_p$, we consider the Gram-Schmidt process described by Murdoch (5).

$$z^*_1 = z'_1$$

$$z^*_k = z'_k - \sum_{j=1}^{k-1} \frac{(z'_k \cdot z^*_j)}{(z^*_j \cdot z^*_j)} z^*_j, \quad k = 2,\dots,p$$

We shall show by induction that $(z^*_i \cdot z^*_j) = 0$ for $i \neq j$.

$$(z^*_1 \cdot z^*_2) = \left(z'_1 \cdot \left[z'_2 - \frac{(z'_2 \cdot z'_1)}{(z'_1 \cdot z'_1)} z'_1\right]\right)$$

$$= (z'_1 \cdot z'_2) - \frac{(z'_2 \cdot z'_1)}{(z'_1 \cdot z'_1)} (z'_1 \cdot z'_1) = 0$$

Now, assume that, for $i \leq m$, $j \leq m$, $i \neq j$, $(z^*_i \cdot z^*_j) = 0$.

Then, for $i \leq m$,

$$(z^*_i \cdot z^*_{m+1}) = z^*_i \cdot \left[z'_{m+1} - \sum_{j=1}^{m} \frac{(z'_{m+1} \cdot z^*_j)}{(z^*_j \cdot z^*_j)} z^*_j\right] =$$

$$= (z^*_i \cdot z'_{m+1}) - \sum_{j=1}^{m} \frac{(z'_{m+1} \cdot z^*_j)}{(z^*_j \cdot z^*_j)} (z^*_i \cdot z^*_j) =$$

$$= (z^*_i \cdot z'_{m+1}) - \frac{(z'_{m+1} \cdot z^*_i)}{(z^*_i \cdot z^*_i)} (z^*_i \cdot z^*_i) = 0$$

Let $z^*_o = z'_o$. We shall show, also by induction, that

$$(z^*_o \cdot z^*_k) = 0 \ , \quad k = 1,\ldots,p$$

$$(z^*_o \cdot z^*_1) = (z'_o \cdot z'_1) = 0 \ .$$

Now assume that for $k = 1,\ldots,m$, $(z^*_o \cdot z^*_k) = 0$. Then

$$(z^*_o \cdot z^*_{m+1}) = \left(z^*_o \cdot \left[z'_{m+1} - \sum_{k=1}^{m} \frac{(z'_{m+1} \cdot z^*_k)}{(z^*_k \cdot z^*_k)} z^*_k\right]\right)$$

$$= (z'_o \cdot z'_{m+1}) - \sum_{k=1}^{m} \frac{(z'_{m+1} \cdot z^*_k)}{(z^*_k \cdot z^*_k)} (z^*_o \cdot z^*_k) = 0$$

Let $z^{(k)} = (z^*_o \ldots z^*_k \quad z'_{k+1} \ldots \ z'_p)$. Let

$$\alpha^{(k)} = \begin{bmatrix} 1 & 0 \ldots & 0 & \ldots 0 \\ 0 & 1 \ldots & -\frac{(z'_k \cdot z^*_1)}{(z^*_1 \cdot z^*_1)} & \ldots\ldots 0 \\ \vdots & \vdots & \vdots & \\ 0 & 0 \ldots & -\frac{(z'_k \cdot z^*_{k-1})}{(z^*_{k-1} \cdot z^*_{k-1})} & \ldots 0 \\ 0 & 0 \ldots & 1 & \ldots 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 \ldots & 0 & \ldots 1 \end{bmatrix}$$

Then $z^{(k)} = z^{(k-1)} \ \alpha^{(k)}$. Thus $z^* = z' \ \alpha^{(2)} \ldots \alpha^{(p)} = z \ \alpha' \ \alpha^{(2)} \ldots \alpha^{(p)}$. Define $\alpha = \alpha' \ \alpha^{(2)} \ldots \alpha^{(p)}$. α is the product of triangular matrices with one's along the diagonal. Therefore α is a triangular matrix with one's along the diagonal.

Finally we remark that, in order for the z^*_i to be well-defined for all $i = 1,2,\ldots p$, it must be the case that $(z^*_i \cdot z^*_i) \neq 0$. Clearly $(z^*_o \cdot z^*_o) = N > 0$. Also, if $(z^*_1 \cdot z^*_1) = 0$, then $z'_1 = z^*_1 = 0$ and thus the vector $(z'_o, z'_1, \ldots, z'_p)$ would be linearly dependent, contrary to assumption. Further, let z^*_i, $i > 1$, be the first vector in the set such that $(z^*_i \cdot z^*_i) \quad 0$ (subsequent vectors would not be defined). Then

$$0 = z^*_i = z'_i + \text{linear combination of } z'_1, \ldots, z'_{i-1}.$$

Again, this is a contradiction of the independence of the vectors $z'_o, z'_1, \ldots, z'_p$.

<u>Corollary</u>. $H = z^T z$ is nonsingular.

Proof: $H = z^T z = (\alpha^{-1})^T z^{*T} z^* \alpha^{-1}$

$$z^{*T}{}^* = \begin{bmatrix} (z^*_o \cdot z^*_o) & (z^*_1 \cdot z^*_o) & \ldots & (z^*_p \cdot z^*_o) \\ (z^*_o \cdot z^*_1) & (z^*_1 \cdot z^*_1) & \ldots & (z^*_p \cdot z^*_o) \\ \vdots & \vdots & & \vdots \\ (z^*_o \cdot z^*_1) & (z^*_1 \cdot z^*_p) & \ldots & (z^*_p \cdot z^*_p) \end{bmatrix} = \begin{bmatrix} (z^*_o \cdot z^*_o) & 0 & \ldots & 0 \\ 0 & (z^*_1 \cdot z^*_1) & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & (z^*_p \cdot z^*_p) \end{bmatrix}$$

But for each $(z^*_i \cdot z^*_i) > 0$. Therefore, $z^{*T}z$ is nonsingular. H is the product of three nonsingular matrices and hence nonsingular.

H is in fact positive definite. In a later chapter, we shall consider this fact and some of its consequences. At present we need only the nonsingularity of H.

<u>Theorem 2</u>. Suppose for each β the model $x^* = z^* \beta^* + \varepsilon$ is the completely orthogonal equivalent of the model $x = z\beta + \varepsilon$, where $z^* = z\alpha$. Then $\beta^* = B^*$ uniquely minimizes $\varepsilon^T \varepsilon$ if and only if $\beta = B = \alpha B^*$ uniquely minimizes $\varepsilon^T\varepsilon$.

Proof: In fact for each β we have $\beta = \alpha\beta^*$. So $\varepsilon^T\varepsilon =$ $(x-z^*\beta^*)^T(x-z^*\beta^*) = (x-z\alpha\beta^*)^T(x-z\alpha\beta^*) = (x-z\beta)^T(x-z\beta)$. Hence if $\beta^* = B^*$ minimizes $\varepsilon^T\varepsilon$, $\beta = B$ must also minimize $\varepsilon^T\varepsilon$, and vice versa.

Since $z^*_o, \ldots, z^*_p$ are $p+1$ linearly independent vectors in an N-dimensional space, there exist $N-p-1$ vectors $\eta_{p+2}, \ldots, \eta_N$ which are linearly independent of the z^*_i and of each other. The Gram-Schmidt process may be applied to the η_j to form N orthogonal vectors $z^*_o, \ldots, z^*_p$, $\eta^*_{p+2}, \ldots, \eta^*_N$.

Let

$$A = \begin{bmatrix} \dfrac{z_o^{*T}}{(z^*_o \cdot z^*_o)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{z_p^{*T}}{(z^*_p \cdot z^*_p)^{\frac{1}{2}}} \\ \dfrac{\eta^{*T}_{p+2}}{(\eta^*_{p+2} \cdot \eta^*_{p+2})^{\frac{1}{2}}} \\ \vdots \\ \dfrac{\eta^{*T}_N}{(\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}}} \end{bmatrix}$$

and consider the linear model

$$y = Ax = Az^* \beta^* + \delta \text{ , where } \delta = A\varepsilon .$$

<u>Theorem 3</u>. $\varepsilon^T\varepsilon$ reaches a minimum at a value if and only if $\delta^T\delta$ does.

Proof: $\delta^T\delta = (A\varepsilon)^T(A\ \ ) = \varepsilon^T\varepsilon A^T A\varepsilon$.

$$AA^T = \begin{bmatrix} \frac{(z^*_o \cdot z^*_o)}{(z^*_o \cdot z^*_o)} & \cdots & \frac{(z^*_o \cdot \eta^*_N)}{(z^*_o \cdot z^*_o)^{\frac{1}{2}} (\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}}} \\ \vdots & & \\ \frac{(\eta^*_N \cdot \ \ ^*_o)}{(\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}} (z^*_o \cdot z^*_o)^{\frac{1}{2}}} & \cdots & \frac{(\eta^*_N \cdot \eta^*_N)}{(\eta^*_N \cdot \eta^*_N)} \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = 1$$

Hence A^T is the inverse of A, i.e., A is orthogonal. It follows that

$$A^TA = I$$

Therefore

$$\delta^T\delta = \varepsilon^T A^T A \varepsilon = \varepsilon^T I \varepsilon = \varepsilon^T\varepsilon$$

Since

$$\delta^T\delta = \varepsilon^T\varepsilon$$

it is trivial that they reach a minimum simultaneously.

<u>Theorem 4</u>. $\delta^T\delta$ reaches a minimum when and only when

$$\beta^*_i = b^*_i = \frac{y_{i+1}}{(z^*_i \cdot z^*_i)^{\frac{1}{2}}} , \quad i = 0,1,\ldots,p$$

Proof: $\delta = y - Az^*\beta^*$

$$Az^* = \begin{bmatrix} \dfrac{z^{*T}_o}{(z^*_o \cdot z^*_o)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{z^{*T}_p}{(z^*_p \cdot z^*_p)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{\eta^{*T}_N}{(\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}}} \end{bmatrix} \begin{bmatrix} z^*_o & \cdots & z^*_p \end{bmatrix}$$

Since the z^*_i and η^*_j are orthogonal

$$Az^* = \begin{bmatrix} (z^*_o \cdot z^*_o)^{\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (z^*_p \cdot z^*_p)^{\frac{1}{2}} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Thus

$$Az^* \beta^* = \begin{bmatrix} (z^*_o \cdot z^*_o)^{\frac{1}{2}} \ \beta^*_o \\ \vdots \\ (z^*_p \cdot z^*_p)^{\frac{1}{2}} \ \beta^*_p \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\delta = y - Az^* \beta^* = \begin{bmatrix} y_1 - (z^*_o \cdot z^*_o)^{\frac{1}{2}} \ \beta^*_o \\ \vdots \\ y_{p+1} - (z^*_p \cdot z^*_p)^{\frac{1}{2}} \ \beta^*_p \\ y_{p+2} \\ \vdots \\ y_N \end{bmatrix}$$

Hence

$$\delta^T \delta = \left[y_1 - (z^*_o \cdot z^*_o)^{\frac{1}{2}} \beta^*_o\right]^2 + \cdots + \left[y_{p+1} - (z^*_p \cdot z^*_p)^{\frac{1}{2}} \beta^*_p\right]^2$$
$$+ y^2_{p+2} + \cdots + y^2_N$$

This is clearly minimized when and only when

$$\beta^*_i = b^*_i = \frac{y_{i+1}}{(z^*_i \cdot z^*_i)^{\frac{1}{2}}}$$

<u>Corollary</u>. $\varepsilon^T\varepsilon$ reaches a minimum when and only when

$$\beta^*_i = b^*_i = \frac{(z^*_i \cdot x)}{(z^*_i \cdot z^*_i)}$$

Proof:

$$y = Ax = \begin{bmatrix} \dfrac{z^{*T}_o}{(z^*_o \cdot z^*_o)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{z^{*T}_p}{(z^*_p \cdot z^*_p)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{\eta^{*T}_N}{(\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}}} \end{bmatrix} x = \begin{bmatrix} \dfrac{(z^*_o \cdot x)}{(z^*_o \cdot z^*_o)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{(z^*_p \cdot x)}{(z^*_p \cdot z^*_p)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{(\eta^*_N \cdot x)}{(\eta^*_N \cdot \eta^*_N)^{\frac{1}{2}}} \end{bmatrix}$$

Thus

$$y_{i+1} = \frac{(z^*_i \cdot x)}{(z^*_o \cdot z^*_i)^{\frac{1}{2}}}, \quad i = 0,\ldots,p$$

$\delta^T\delta$ is minimized when and only when

$$\beta^*_i = b_i = \frac{y_{i+1}}{(z^*_i \cdot z^*_i)^{\frac{1}{2}}} = \frac{(z^*_i \cdot x)}{(z^*_i \cdot z^*_i)}$$

Since $\varepsilon^T\varepsilon = \delta^T\delta$, $\varepsilon^T\varepsilon$ is minimized when and only when

$$\beta^*_i = b^*_i = \frac{(z^*_i \cdot x)}{(z^*_i \cdot z^*_i)}$$

Theorem 5. $B = \alpha B^*$ is the unique solution to the normal equations

$$HB = G$$

Proof: Consider the system of equations

$$HB = G$$

where

$$H = z^T z \ , \quad G = z^T x$$

That is,

$$z^T z B = z^T x$$

This system of equations is equivalent to

$$(\alpha^{-1})^T z^{*T} z^* \alpha^{-1} B = (\alpha^{-1})^T (z^*)^T x$$

or, since $(\alpha^{-1})^T$ is nonsingular,

$$z^{*T} z^* \alpha^{-1} B = z^{*T} x \ .$$

Now, from the corollary to Theorems 3 and 4, it is seen that

$$(z^*_i \cdot z^*_i)\, b^*_i = (z^*_i \cdot x)$$

or

$$z^{*T} z^* B^* = z^{*T} x$$

Hence $B = \alpha B^*$ is a solution to $HB = G$. However, since H is nonsingular, there is a unique solution. Therefore, $B = \alpha B^*$ is the unique solution to $HB = G$.

Theorem 6. If in the models

$$x = z\beta + \varepsilon,$$

where β is a parameter vector, the vectors z_i are linearly independent, then the unique β which minimizes $\varepsilon^T\varepsilon$ and thus provides the best fitting plane in the sense of least squares is $\beta = B$, which satisfies uniquely the normal equations $HB = G$.

Proof: By Theorem 1, there exists a completely orthogonal equivalent linear model.

$$x = z^* \beta^* + \varepsilon$$

By the corollary to Theorems 3 and 4 $\varepsilon^T\varepsilon$ is uniquely minimized by the solution B^* to the equivalent normal equations

$$H^*B^* = G^* .$$

By Theorem 2 $\varepsilon^T\varepsilon$ is also uniquely minimized by $B = \alpha B^*$. By Theorem 5 $B = \alpha B^*$ is the unique solution of the normal equations $HB = G$. Thus, $\varepsilon^T\varepsilon$ is uniquely minimized by the unique solution of the normal equations

$$HB = G .$$

In the next chapter it will be shown that, if we return to the statistical regression model (1.1), where ε_μ are random variables and the β_i are unknown parameters, then the problem of finding unbiased minimum variance linear estimates of the β_i formally reduces to solving the normal equations, $HB = G$, above.

CHAPTER III

STATISTICAL ESTIMATION OF THE PARAMETERS IN THE REGRESSION MODEL

In the last chapter we determined a plane for approximating a "dependent" variable $\tilde{x}$ as a linear function of p "independent" variables $\tilde{z}_1,\ldots,\tilde{z}_p$, which was "best-fitting" to a set of N observed points $(x_\mu, z_{\mu 1},\ldots,z_{\mu p})$ in the sense of least squares -- that is, in the sense of giving a "prediction" formula

$$\hat{x} = b_o + \sum_{i=1}^{p} b_i \tilde{z}_i$$

such that $\sum_{u=1}^{N} (x_\mu - \hat{x}_\mu)^2$ was the least that could be obtained for any set of coefficients.

In this chapter we shall revert to the statistical model, described in Chapter I as the regression model, and show that the optimal estimation of the parameters β_o, $\beta_1,\ldots,\beta_p$ in that model, from a given set of statistical data (the random and structural vectors), turns out to be formally identical with solving the normal equations in the least squares problem of Chapter II. In addition certain useful formulas will be derived. Recall the regression model (1.1b)

$$x = z\beta + \varepsilon$$

Defining the expected value of a matrix as the matrix of expected values, in the regression model the random vector x had the property $Ex = z\beta$, where z was a matrix of structural, nonrandom vectors and β was a vector of unknown parameters. We might conceive of the model arising from a sampling from the conditional distribution of x, given various sets of values of the vector $(z_1,\ldots,z_p)$, where we are assuming that the regression surface for the mean of x is linear; or we might conceive of the ε_μ as random deviations of the random observations x_μ from the mean square regression plane of x on $z_1,\ldots,z_p$ at the various points $(z_{\mu 1},\ldots z_{\mu p})$, where the regression plane is defined by the unknown parameters β_i.

<u>Lemma 1</u>. For a random vector x and a nonrandom matrix A, $Eax = AEx$.

Proof: If $A = (a_{ij})$, $x = (x_j)$

$$E(Ax) = \left(E\left(\sum_j a_{ij}\ x_j\right)\right) = \left(\sum_j a_{ij}\ Ex_j\right) = AEx$$

<u>Lemma 2</u>. If K, L are matrices of nonrandom numbers and A is a matrix of random variables, then

$$E(KAL) = K(EA)L$$

Proof: If $K = (k_{hi})$, $A = (a_{ij})$, $L = (\ell_{jk})$

$$E(KAL) = \left(E\left(\sum_i \sum_j k_{hi}\ a_{ij}\ \ \ell_{jk}\right)\right) = \left(\sum_i \sum_j E(K_{hi} a_{ij}\ \ell_{jk})\right) =$$

$$= \left(\sum_i \sum_j k_{hi}(Ea_{ij})\ \ell_{jk}\right) = K(EA)L$$

<u>Lemma 3</u>. Let x' be a random vector. Let $x = x' - Ex'$. If $x = Ay$, where A is nonrandom, then

$$Exx^T = A(Eyy^T)\ A^T$$

Proof:

$$xx^T = (Ay)(Ay)^T = Ayy^T A^T$$

Therefore, by Lemma 2,

$$Exx^T = E(Ayy^T A^T) = A(Eyy^T) A^T$$

Theorem 1. If $x = z\beta + \varepsilon$ where $E\varepsilon = 0$, and if B is the solution to the normal equations $HB = G$, then $EB = \beta$; i.e., the solutions B of the normal equations are unbiased estimates of β.

Proof: The solution of $HB = G$ is $B = H^{-1} G$.

$$EB = E(H^{-1}G) = H^{-1} EG$$

$$EG = E(z^T x) = z^T Ex$$

$$Ex = E(z\beta + \varepsilon) = Ez\beta + E\varepsilon = z\beta$$

Hence

$$EG = z^T z\beta = H\beta$$

Therefore

$$EB = H^{-1} H\beta = \beta$$

The covariance matrix of the vectors $x_1, \ldots, x_n$ is $E(x-Ex)(x-Ex)^T$, where

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Theorem 2. The covariance matrix of the solution vector B of the normal equations is

$$E(B - EB)(B - EB)^T = H^{-1} z^T E\varepsilon\varepsilon^T zH^{-1}$$

Proof:

$$B - EB = H^{-1} G - EH^{-1} G = H^{-1} (G - EG)$$

$$E(B - EB)(B - EB)^T = EH^{-1} (G - EG)(G - EG)^T (H^{-1})^T =$$

$$= H^{-1} E(G-EG)(G-EG)^T(H^{-1})^T = H^{-1}E(G-EG)(G-EG)^T H^{-1} ,$$

since H and H^{-1} are symmetric. Now

$$G - EG = z^T x - Ez^T x = z^T (x - Ex)$$

$$E(G - EG)(G - EG)^T = Ez^T(x - Ex)(x - Ex)^T z =$$

$$= z^T E(x - Ex)(x - Ex)^T z$$

But

$$x - Ex = \varepsilon$$

and therefore

$$E(x - Ex)(x - Ex)^T = E\varepsilon\varepsilon^T$$

Hence

$$E(G - EG)(G-EG)^T = z^T E\varepsilon\varepsilon^T z$$

Therefore

$$E(B - EB)(B - EB)^T = H^{-1} z^T E\varepsilon\varepsilon^T zH^{-1}$$

<u>Corollary 1</u>. If the ε_μ are uncorrelated with $E\varepsilon_\mu^2 = \sigma_\mu^2$, and if $\sum = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix}$, then the covariance matrix of the b_i is

$$E(B - EB)(B - EB)^T = H^{-1} z^T \sum zH^{-1}$$

Proof: Since the ε_μ are uncorrelated, for $\mu \neq \nu$

$$E\varepsilon_\mu\varepsilon_\nu = E\varepsilon_\mu E\varepsilon_\nu = 0 .$$

$$E\varepsilon\varepsilon^T = (E\varepsilon_\mu \varepsilon_\nu) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix} = \sum$$

$$E(B - EB)(B - EB)^T = H^{-1} z^T E\varepsilon\varepsilon^T zH^{-1}$$

Therefore

$$E(B - EB)(B - EB)^T = H^{-1} z^T \sum zH^{-1}$$

Corollary 2. If the ε_μ are uncorrelated with common variance σ^2, then the covariance matrix of the b_i is

$$E(B - EB)(B - EB)^T = \sigma^2 H^{-1}$$

Proof:

$$\sum = \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Hence

$$E(B-EB)(B-EB)^T = H^{-1}z^T(\sigma^2 I)zH^{-1} = \sigma^2 H^{-1} z^T IzH^{-1}$$

$$= \sigma^2 H^{-1} z^T zH^{-1} = \sigma^2 H^{-1}HH^{-1} = \sigma^2 H^{-1}$$

Theorem 3 (Gauss-Markov). Consider the regression model

$$x = z\beta + \varepsilon$$

where $E\varepsilon = 0$ and $E\varepsilon\varepsilon^T = \sigma^2 I$. Let $\tilde{x} = \beta_o \tilde{z}_o + \cdots + \beta_p \tilde{z}_p$ be any linear combination of $\beta_o, \ldots, \beta_p$ with $\tilde{z}_o = 1$. Again write

$$B = \begin{bmatrix} b_o \\ \vdots \\ b_p \end{bmatrix} = H^{-1}G, \text{ where } H = z^T z \text{ and } G = z^T x,$$

the solution vector of the normal equations. Then

$$\hat{x} = b_o + \sum_{i=1}^{p} b_i \tilde{z}_i$$

is an unbiased estimate of $\tilde{x}$, and $\tilde{x}$ is linear in the random variables $x_1, x_2, \ldots, x_N$. Furthermore, in the class of all such linear unbiased estimates of $\tilde{x}$, $\hat{x}$ has minimum variance.

Proof: Write $\tilde{z} = \begin{bmatrix} z_o \\ \vdots \\ z_p \end{bmatrix}$. Then $\tilde{x} = \tilde{z}^T \beta$, $\hat{x} = \tilde{z}^T B$

$$E\hat{x} = E\tilde{z}^T B = \tilde{z}^T EB = \tilde{z}^T \beta = \tilde{x} ,$$

by Theorem 1. Thus $\hat{x} = \tilde{z}^T B$ is an unbiased estimate of $\tilde{x} = \tilde{z}^T$. Also,

$$\hat{x} = \tilde{z}^T B = \tilde{z}^T H^{-1} G = (\tilde{z}^T H^{-1} z^T) x ,$$

from which it is clear that $\hat{x}$ is a linear function of $x_1, \ldots, x_N$.

Following the proof of Scheffe (6), p. 14, we shall now show that $\hat{x}$ has minimum variance in the class of unbiased linear estimates of $\tilde{x}$.

Since $\hat{x}$ is an unbiased linear estimate of $\tilde{x}$, there exists at least one vector a in N-dimensional space such that

$$E(a^T x) = \tilde{x} .$$

Consider such a vector a. Recall from Chapter II the orthogonal transform of z obtained by the Gram-Schmidt method:

$$z^* = z\alpha .$$

Normalizing the vectors of this matrix, we obtain

$$w = z^* \gamma = z \alpha \gamma ,$$

where γ is a diagonal matrix with $(z^*_i \cdot z^*_i)^{-\frac{1}{2}}$ in the i^{th} diagonal position.

Consider the vector

$$a^* = wc ,$$

where

$$c^T = a^T w .$$

Clearly

$$a^* = wc = z(\alpha \gamma w^T a)$$

a linear combination of the vector $z_0, z_1, \ldots, z_p$. Now

$$a^{*T} w = c^T w^T w = c^T ,$$

since $w^T w = I$. But $c^T = a^T w$, and therefore

$$(a^{*T} - a^T) w = 0.$$

It follows that

$$(a^{*T} - a^T) z = 0 \tag{3.1}$$

Next

$$E(a^T x) = E(a^{*T} x) + E\left((a-a^*)^T x\right) = E(a^{*T} x) + (a-a^*)^T z\beta$$

However, since $(a-a^*)^T z = 0$,

$$E(a^T x) = E(a^{*T} x) = \tilde{x} .$$

Now we shall show that a^* is the only vector lying in the vector space generated by $z_0, z_1, \ldots, z_p$ such that $Ea^{*T}x = \tilde{x}$. Let $E(\alpha^{*T} x) = \tilde{x}$ be a linear estimate with $\alpha^* = zd'$. Write $d = \alpha\gamma c$, so that $a^* = zd$. Then

$$0 = E(a^* - \alpha^*)^T x = (a^* - \alpha^*)^T Ex = (a^* - \alpha^*)^T z\beta =$$

$$= (zd - zd')^T z\beta = (d-d)^T z^T z\beta = (d-d')^T H\beta$$

Since this is true regardless of the values of β, $(d-d')^T H = 0$. Since H is nonsingular, $d-d' = 0$. Thus $a^* - \alpha^* = 0$.

From equation (3.1) it is evident that a^* and $a - a^*$ are orthogonal, since a^* is a linear combination of the z_i. It is well-known that the representation of a as the sum of two vectors, one lying in a subspace and the other orthogonal to it, is unique. The vector a^* is called the orthogonal projection of a onto the space generated by $z_0, z_1, \ldots, z_p$. We have thus shown that there exists one and only one unbiased linear estimate $a^{*T}x$, of x such that a^* lies in the subspace spanned by $\{z_i\}$, and also we have shown that the projection of every unbiased linear estimate of $\tilde{x}$ onto this subspace is therefore necessarily a^*.

It is evident that $a^{*T}x = \hat{x}$. For

$$\hat{x} = (zH^{-1}\tilde{z})^T x ,$$

where $zH^{-1}\tilde{z}$ is a vector lying in the space of $\{z_i\}$. As we have already seen, $\hat{x}$ is unbiased. Therefore, from the uniqueness property above, $z^* = zH^{-1}\tilde{z}$.

We shall now show that, in the class of unbiased linear estimates of $\tilde{x}$, $\hat{x} = a^{*T} x$ has minimum variance. Note that

$$a^T a = a^{*T} a^* + a^{*T}(a - a^*) + (a-a^*)^T a + (a-a^*)^T (a-a^*)$$

However

$$(a - a^*)^T a^* = (a - a^*)^T zd = 0$$

Therefore

$$a^T a = a^{*T} a^* + (a - a^*)^T (a - a^*)$$

The variance of $a^T x$ is $E(a^T x - Ea^T x)(a^T x - Ea^T x)^T$. Now

$$a^T x - Ea^T x = a^T(x - Ex) = a^T(x - z\beta) = a^T \varepsilon$$

Thus

$$\text{var } (a^T x) = Ea^T \varepsilon\varepsilon^T a = a^T(E\varepsilon\varepsilon^T) a$$

Since $E\varepsilon\varepsilon^T = \sigma^2 I$,

$$\text{var } (a^T x) = \sigma^2 a^T a = \sigma^2 a^{*T} a^* + \sigma^2 (a-a^*)^T (a-a^*)$$

$$= \text{var } (a^{*T} x) + \sigma^2 (a-a^*)^T (a-a^*)$$

Thus var $(a^{*T}x) \leq$ var $(a^T x)$ with equality if and only if $a = a^*$. Hence $a^{*T}x$ is the unique unbiased linear estimate of $\tilde{\gamma}$ with minimum variance.

<u>Theorem 4</u>. If in the linear model

$$x = z\beta + \varepsilon$$

the ε_μ are $N(0, \sigma^2)$ and independent, then the maximum likelihood estimates of β are the solutions to $HB = G$.

Proof: The method of maximum likelihood consists in choosing those values of the parameters $\beta_0, \beta_1, \ldots, \beta_p$ which will maximize the so-called likelihood function, i.e., the joint density function in this case of the ε_μ. (See Cramer (2), Chapter 33, for optimal characteristics of maximum likelihood estimates.)

But, under the hypotheses, this density function is

$$\frac{1}{(\sigma^2 2\pi)^{N/2}} e^{-\frac{1}{2\sigma^2} S\varepsilon_\mu^{\ 2}}$$

which is clearly maximized when $S\varepsilon_\mu^{\ 2}$ is minimized. However, we have already seen in Chapter II that $S\varepsilon_\mu^{\ 2}$ is uniquely minimized by $B = H^{-1}G$.

<u>Definition 2</u>. In the regression model, a <u>residual</u> is defined as

$$e_\mu = x_\mu - \hat{x}_\mu$$

where

$$\hat{x}_\mu = \sum_{i=0}^{p} b_i z_{\mu i}$$

<u>Theorem 5</u>. In the regression model the sum of squares of residuals is

$$SSE \equiv e^T e = x^T x - B^T G = Sx_\mu^{\ 2} - N\bar{x}^2 - \sum_{i=0}^{p} b_i g_i ,$$

where we write $e^T = (e_1, \ldots, e_N)$

Proof:

$$e^T e = (x-aB)^T(x-zB) = x^T x - x^T zB - B^T z^T x - B^T z^T zB$$

$$= x^T x - G^T B - B^T G - B^T HB = x^T x - 2B^T G - B^T HB$$

However, B is a solution of $HB = G$. Thus we have $e^T e = x^T x - B^T G$ or, in summation notation

$$Se_\mu^{\ 2} = Sx_\mu^{\ 2} - \sum_{i=0}^{p} b_i g_i .$$

CHAPTER IV

TESTING HYPOTHESES AND ANALYSIS OF VARIANCE

In the last chapter, in the Gauss-Markov theorem, it was shown that if in the regression model

$$x = z\beta + \varepsilon$$

$E\varepsilon = 0$, $E\varepsilon_\mu \varepsilon_\nu = \sigma^2 \delta_{\mu\nu}$, where $\delta_{\mu\nu}$ is the Kronecker delta, if $\tilde{x} = \beta_0 + \sum_{i=1}^{p} \beta_i \tilde{z}_i$, a linear function of the β_i, then of all the unbiased estimates of $\tilde{x}$, linear in $\tilde{x}$, the one of minimum variance is

$$\hat{x} = \sum_{i=1}^{p} b_i \tilde{z}_i, \quad \text{where } B = H^{-1} G .$$

For the purpose of constructing tests of hypotheses and making use of extant normal theory, in addition to the hypotheses stated in the Gauss-Markov theorem, we now impose a further hypothesis on the regression model; we assume that the ε_μ are normal random variables. Thus ε_μ is $N(0, \sigma^2)$ and independent for $\mu = 1, \ldots, N$.

The pertinent distributions, normal, chi-square, t, and F, are developed in the appendix.

The general problem we wish to solve is how to test the hypothesis

$$H_0: \beta_{r+1} = \cdots = \beta_p = 0$$

for any r.

Recall in the completely orthogonal equivalent model

$$x = z^* \beta^* + \varepsilon$$

that

$$\beta = \alpha \beta^* \text{ , where } \alpha = \begin{bmatrix} 1 & \alpha_{01} & \cdots & \alpha_{0p} \\ 0 & 1 & \cdots & \alpha_{1p} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{4.1}$$

<u>Theorem 1</u>. For any $r = 0, 1, \ldots, p-1$,

$$\beta_{r+1} = \cdots = \beta_p = 0 \quad \text{if and only if}$$
$$\beta^*_{r+1} = \cdots = \beta^*_p = 0$$

Write

$$\beta(r) = \begin{bmatrix} \beta_{r+1} \\ \vdots \\ \beta_p \end{bmatrix} , \quad \beta^*(r) = \begin{bmatrix} \beta^*_{r+1} \\ \vdots \\ \beta^*_p \end{bmatrix} , \quad \alpha(r) = \begin{bmatrix} 1 & \alpha_{r+1,r+2} & \cdots & \alpha_{r+1,p} \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

Clearly, from (4.1),

$$\beta(r) = \alpha(r) \ \beta^*(r),$$

whence it follows that $\beta(r) = 0$ if and only if $\beta^*(r) = 0$, since $\alpha(r)$ is nonsingular.

Note that from Theorem 1, Chapter III,

$$EB^* = \beta^*$$

where B^* is the solution of $H^*B^* = G^*$.

Since the ε_μ are $N(0, \sigma^2)$, the $x_\mu = \sum_{i=0}^{r} \beta_i z_i + \varepsilon_\mu$ are $N\left(\sum_{i=0}^{p} \beta_i z_{\mu i}, \sigma^2\right)$ variables.

Before making the orthogonal transformation $y = Ax$, defined in Chapter II, we shall prove the following fundamental theorem.

<u>Theorem 2.</u> If $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$ is a vector of $N(m_\mu, \sigma^2)$ independent variables, and $y = Ax$, where $A = (a_{\mu\nu})$ is orthogonal, then the y_μ are $N\left(\sum_{\nu=1}^{N} a_{\mu\nu} m_\nu, \sigma^2\right)$ independent variables.

Proof: The joint distribution of the x_μ is

$$F(x) = P\{\xi : \xi \le x\} = \int_{-\infty}^{x} (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{(\xi - m)^T(\xi - m)}{2\sigma^2}\right\} d\xi$$

The distribution of the y_μ is

$$G(y) = P\{\xi : A\xi \le y\} = P\{\xi : \xi \le A^{-1}y\} = F(A^{-1}y) =$$

$$= \int_{-\infty}^{A^{-1}y} (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{(\xi - m)^T(\xi - m)}{2\sigma^2}\right\} d\xi$$

We make the change of variable $\eta = A\xi$. Then

$$G(y) = \int_{-\infty}^{y} (2\pi\sigma^2)^{-N/2} \exp\{-(A^{-1}\eta - m)^T(A^{-1}\eta - m)/2\sigma^2\} \,|\det A^{-1}|\, d\eta$$

Since A is orthogonal, $|A| = \pm 1$ and

$$(A^{-1}\eta - m)^T(A^{-1}\eta - m) = (\eta - Am)^T(A^{-1})^T A^{-1}(\eta - Am) = (\eta - Am)^T A A^{-1}(\eta - Am)$$

$$= (\eta - Am)^T(\eta - Am)$$

Hence

$$G(y) = \int_{-\infty}^{y} (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{(\eta - Am)^T (\eta - Am)}{2\sigma^2}\right) d\eta$$

Thus the y_μ are $N\left(\sum_{\nu=1}^{N} a_{\mu\nu} m_\nu ,\ \sigma^2\right)$ independent.

<u>Corollary 1</u>. If the x_μ are $N(0, \sigma^2)$ independent and $y = Ax$, where A is orthogonal, then the y_μ are $N(0, \sigma^2)$ independent.

Proof: Let $m = 0$, and apply theorem 2.

<u>Corollary 2</u>. In the regression model

$$x = z^* \beta^*$$

if the vector x is transformed orthogonally by $y = Ax$, then writing $\delta = A\varepsilon$, the δ_μ are $N(0, \sigma^2)$ independent.

Proof: Corollary 1 is applied to $\delta = A\varepsilon$.

<u>Corollary 3</u>. If $x = z^* \beta^* + \varepsilon$ and $y = Ax$, where A is orthogonal, then

$$Ey_{i+1} = \beta^*_i \,(z^*_i \cdot z^*_i)^{\frac{1}{2}} , \text{ for } i = 0,1,\ldots,p$$

$$Ey_\mu = 0, \qquad \text{for} \qquad \mu = p + 2, \ldots, N$$

Proof: $y = Ax = A(z^* \beta^* + \varepsilon) = Az^* \beta^* + \delta$

From Chapter II,

$$A = \begin{bmatrix} \dfrac{z^*_o}{(z^*_o \cdot z^*_o)^{\frac{1}{2}}} \\ \vdots \\ \dfrac{z^*_p}{(z^*_p \cdot z^*_p)^{\frac{1}{2}}} \\ \\ \text{N-p-1 other} \\ \text{orthog vectors} \end{bmatrix}$$

Hence

$$Az^* \beta^* = \begin{bmatrix} (z^*_o \cdot z^*_o)^{\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & (z^*_1 \cdot z^*_1)^{\frac{1}{2}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (z^*_p \cdot z^*_p)^{\frac{1}{2}} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \beta^*_o \\ \beta^*_1 \\ \vdots \\ \beta^*_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \beta^*_o \; (z^*_o \cdot z^*_o)^{\frac{1}{2}} \\ \beta^*_1 \; (z^*_1 \cdot z^*_1)^{\frac{1}{2}} \\ \vdots \\ \beta^*_p \; (z^*_p \cdot z^*_p)^{\frac{1}{2}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Thus

$$y = \begin{bmatrix} \beta^*_o \; (z^*_o \cdot z^*_o)^{\frac{1}{2}} \\ \vdots \\ \beta^*_p \; (z^*_p \cdot z^*_p)^{\frac{1}{2}} \\ \vdots \\ 0 \end{bmatrix} + \delta$$

or

$$y_{i+1} = \beta^*_i \; (z^*_i \cdot z^*_i)^{\frac{1}{2}} \quad \delta_i, \quad i = 0,1,\ldots,p$$

$$y_\mu = \delta_\mu \quad , \qquad \mu = p + 2, \ldots, N \; .$$

Since $E\delta_\mu = 0$,

$$Ey_{i+1} = \beta^*_i (z^*_i \cdot z^*_i)^{\frac{1}{2}}, \quad i = 0,1,\ldots,p$$

$$Ey_\mu = 0 \quad , \qquad \mu = p+2,\ldots,N$$

Now we write

$$\gamma_i = \beta^*_i (z^*_i \cdot z^*_i)^{\frac{1}{2}} \qquad i = 0,1,\ldots,p$$

By Theorem 1 H_o is equivalent to $\beta^*_{r+1} = \cdots = \beta^*_p = 0$, and since $(z^*_i \cdot z^*_i) > 0$, H_o is equivalent to $\gamma_{r+1} = \cdots = \gamma_p = 0$.

Writing $\gamma = \begin{bmatrix} \gamma_o \\ \vdots \\ \gamma_p \end{bmatrix}$, $C = \begin{bmatrix} c_{oo} & \cdots & c_{op} \\ \vdots & & \vdots \\ c_{po} & \cdots & c_{pp} \end{bmatrix} = \begin{bmatrix} (z^*_o \cdot z^*_o)^{\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & (z^*_p \cdot z^*_p)^{\frac{1}{2}} \end{bmatrix}^{-1} \cdot \alpha^{-1}$

Clearly $\gamma = C\beta$. Thus we arrive at the situation where we have N normal independent variables y_μ with common variance such that

	Under general hypothesis	Under H_o
$Ey_1 =$	γ_o	γ_o
$\vdots$	$\vdots$	$\vdots$
$Ey_{r+1} =$	γ_r	γ_r
$Ey_{r+2} =$	γ_{r+1}	0
$\vdots$	$\vdots$	$\vdots$
$Ey_{p+1} =$	γ_p	0
$Ey_{p+2} =$	0	0
$\vdots$	$\vdots$	$\vdots$
$Ey_N =$	0	0

Regarding the general problem of testing a hypothesis concerning a set of parameters:

Let $f(x;\,\theta)$ denote the family of joint probability density functions of a vector x depending on a vector θ of parameters, where θ lies in a parameter space Ω. Let the composite hypothesis to be tested be

$$H_0: \quad \theta \varepsilon \omega$$

against the alternative that

$$H_1: \theta \varepsilon \Omega - \omega$$

Let C be a <u>critical region</u> of points in the sample space such that, if an observed sample is drawn from this region, the decision is made to reject H_0 (accept H_1). It is agreed that the <u>size</u> of the critical region will be fixed by what chance of Type I error (rejecting H_0 when it is true) the experimenter is willing to tolerate. Thus we may require

$$\int_{\omega} f(x;\, \theta) \quad dx \leq \varepsilon \qquad \text{for every } \theta \varepsilon \omega$$

that is, the probability of rejecting H_0 when it is true shall not exceed a fixed ε. A desirable feature of any test of given size is to make it as powerful as possible - that is, of all tests of given size, we should like to choose one which gives greatest assurance that the Type II error (accepting H_0 when it is false) will be least in some sense.

The likelihood ratio method for determining a critical region runs briefly as outlined below. For a discussion of the optimal character of tests based on this parameter, reference is made to Wald (7).

$$\text{Let } f(x;\, \hat{\hat{\theta}}) = \max_{\theta \varepsilon \omega} f(x;\, \theta)$$

$$f(x;\, \hat{\theta}) = \max_{\theta \varepsilon \omega} f(x;\, \theta)$$

$$\lambda = \lambda(x) = \frac{f(x;\, \hat{\hat{\theta}})}{f(x;\, \hat{\theta})}$$

Obviously $\lambda \geq 1$, since ω is contained in Ω, and therefore the maximum over ω cannot be greater than the maximum over Ω.

Define $C(k) = \{x: \lambda \geq k\}$, for $k \geq 1$. We then try to choose the largest value of k so that under H_o

$$P_\theta \{C(k)\} \leq \varepsilon$$

We shall call this value of k, if it exists, λ_ε, and write the critical region

$$C = C(\lambda_\varepsilon) = \{x: \lambda(x) \geq \lambda_\varepsilon\}$$

In the present instance we shall without difficulty be able to find the critical value λ_ε. The region Ω is the collection of all vectors

$$\theta = \begin{bmatrix} \gamma_o \\ \vdots \\ \gamma_p \\ \sigma^2 \end{bmatrix}.$$

The region ω is the subset of Ω where

$$\gamma_{r+1} = \cdots = \gamma_p = 0$$

The likelihood function relative to the distribution of y is

$$f(y,\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{p+1} (y_i - \gamma_{i-1})^2 + \sum_{\mu=p+2}^{N} y_\mu^2\right]$$

$$\log f = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{p+1} (y_i - \gamma_{i-1})^2 + \sum_{\mu=p+2}^{N} y_\mu^2\right]$$

$$\frac{\partial \log f}{\partial \gamma_i} = \frac{y_{i+1} - \gamma_i}{\sigma^2} = 0, \quad i = 0, .,\ldots, p$$

$$\frac{\partial \log f}{\partial \sigma^2} = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{(\sigma^2)^2}\left[\sum_{i=1}^{p+1} (y_i - \gamma_{i-1})^2 + \sum_{\mu=p+2}^{N} y_\mu^2\right] = 0$$

Thus the maximum likelihood estimates over Ω are $\hat{\hat{\gamma}}_i = y_{i+1}$

$$\hat{\hat{\sigma}}^2 = \frac{2}{N} \sum_{\mu=p+2}^{N} y_\mu^2$$

Thus $$f(x;\, \hat{\hat{\theta}}) = \frac{1}{\left[2\pi e \sum_{\mu=p+2}^{N} y_\mu^2\right]^{N/2}}$$, the exponential reducing to e^{-1}.

Similarly, the maximum likelihood estimates over ω are

$$\hat{\gamma}_i = y_{i+1} \qquad i = 0, \ldots, r$$

$$\hat{\sigma}^2 = \sum_{\mu=r+2}^{N} y_\mu^2$$

and

$$f(x;\, \hat{\theta}) = \frac{1}{\left[2\pi \sum_{\mu=r+2}^{N} y_\mu^2\right]^{N/2}}$$

Hence

$$\lambda = \frac{f(x;\, \hat{\hat{\theta}})}{f(x,\, \hat{\theta})} = \left[\frac{\sum_{\mu=r+2}^{N} y_\mu^2}{\sum_{\mu=p+2}^{N} y_\mu^2}\right]^{N/2} = 1 + \left[\frac{\sum_{i=r+2}^{p+1} y_i^2}{\sum_{\mu=p+2}^{N} y_\mu^2}\right]^{N/2}$$

Thus

$$C(k) = \{x:\ \lambda(x) \geq k\} =$$

$$= \left\{ x:\ \left[1 + \frac{\sum_{i=r+2}^{p+1} y_i^2}{\sum_{\mu=p+2}^{-N} y_\mu^2}\right]^{N/2} \geq k \right\} =$$

$$= \left\{ x : \frac{\sum_{i=r+2}^{p+1} y_i^2 \Big/ (p-r)}{\sum_{\mu=p+2}^{N} y_\mu^2 \Big/ (N-p-1)} \geq (k^{2/N} - 1) \frac{N-p-1}{p-r} = F_k \right\}$$

Now under H_o the statistic

$$F_{p-r,N-p-1} = \frac{\sum_{i=r+2}^{p+1} y_i^2 \Big/ (p-r)}{\sum_{\mu=p+2}^{N} y_\mu^2 \Big/ (N-p-1)}$$ is distributed in an F distribution with

p-r and N-p-1 degrees of freedom (see Appendix). It is thus an easy matter to determine F_o, and therefore λ_ε, such that under H_o (when $\theta \in \omega$)

$$P_\theta\{\lambda(x) \geq \lambda_\varepsilon\} = P\{F_{p-r,N-p-1} \geq F_o\} = \varepsilon$$

This testing procedure is summarized in Table 1, which may be condensed to form Table 2.

The test statistic to be computed is

$$F_{p-r,N-p-1} = \frac{\sum_{i=1}^{p+1} y_i^2 \Big/ (p-r)}{\sum_{\mu=p+2}^{N} y_\mu^2 \Big/ (N-p-1)}$$

This statistic is then compared with critical values from standard F tables, such as those given by Fraser (3) or Mood (4).

Table 1 may be translated into terms of the original regression problem. Table 1A gives the results. It may be condensed into Table 2A.

Table 1

Analysis of Variance Table in Terms of Orthonormal Variables

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Sum of Squares (MS)	Expected Value of MS (EMS) General	H_o
γ_o ⋮ γ_r	y_1^2 ⋮ y_{r+1}^2	1 ⋮ 1	y_1^2 ⋮ y_r^2	$\sigma^2 + \gamma_o^2$ ⋮ $\sigma^2 + \gamma_r^2$	$\sigma^2 + \gamma_o^2$ ⋮ $\sigma^2 + \gamma_r^2$
γ_{r+1} ⋮ γ_p	y_{r+2}^2 ⋮ y_{p+1}^2	1 ⋮ 1	y_{r+2}^2 ⋮ y_p^2	$\sigma^2 + \gamma_{r+1}^2$ ⋮ $\sigma^2 + \gamma_p^2$	σ^2 ⋮ σ^2
Error	$\sum_{\mu=p+2}^{N} y_\mu^2$	N-p-1	$\sum_{\mu=p+2}^{N} \frac{y_\mu^2}{N-p-1}$	σ^2	σ^2
Total	Sy_μ^2	N			

Table 2

Condensed Analysis of Variance
Table in Terms of Orthonormal Variables

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Sum of Squares (MS)	Expected Value of MS: General	Expected Value of MS: H_o
$\gamma_o, \ldots, \gamma_r$	$\sum_{i=1}^{r+1} y_i^2$	r	$\sum_{i=1}^{r+1} \frac{y_i^2}{r}$	$\sigma^2 \sum_{i=1}^{r+1} \frac{\gamma_i^2}{r}$	$\sigma^2 \sum_{i=1}^{r+1} \frac{\gamma_i^2}{r}$
$\gamma_{r+1}, \ldots, \gamma_p$	$\sum_{i=r+2}^{p+1} y_i^2$	p-r	$\sum_{i=r+2}^{p+1} \frac{y_i^2}{p-r}$	$\sigma^2 \sum_{i=r+2}^{p+1} \frac{\gamma_i^2}{p-r}$	σ^2
Error	$\sum_{\mu=p+2}^{N} y_\mu^2$	N-p-1	$\sum_{\mu=p+2}^{N} \frac{y_\mu^2}{N-p-1}$	σ^2	σ^2
Total	Sy_μ^2	N			

To test γ_o: $\gamma_{r+1} = \ldots = \gamma_p = 0$, compute

$$\frac{\sum_{i=r+2}^{p+1} y_i^2 \Big/ (p-r)}{\sum_{\mu=p+2}^{N} y_\mu^2 \Big/ (N-p-1)}$$

and reject H_o if this exceeds the critical value given in the F Tables.

Table 1A

Analysis of Variance Table in Terms of β^*-Weights

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Sum of Squares (MS)	Expected Value of MS	
				General	H_o
$\gamma_o = \beta^*_o (z^*_o \cdot z_o)^{\frac{1}{2}}$ $\vdots$ $\gamma_r = \beta^*_r (z^*_r \cdot z^*_r)^{\frac{1}{2}}$	$b^{*2}_o (z^*_o \cdot z^*_o)$ $\vdots$ $b^{*2}_r (z^*_r \cdot z^*_r)$	1 $\vdots$ 1	$b^{*2}_o (z^*_o \cdot z^*_o)$ $\vdots$ $b^{*2}_r (z^*_r \cdot z^*_r)$	$\sigma^2 + \beta^{*2}_o (z^*_o \cdot z^*_o)$ $\vdots$ $\sigma^2 + \beta^{*2}_r (z^*_p \cdot z^*_r)$	$\sigma^2 + \beta^{*2}_o (z^*_o \cdot z^*_o)$ $\vdots$ $\sigma^2 + \beta^{*2}_r (z^*_r \cdot z^*_r)$
$\gamma_{r+1} = \beta^*_{r+1} (z^*_{r+1} \cdot z^*_{r+1})$ $\vdots$ $\gamma_p = \beta^*_p (z^*_p \cdot z^*_p)$	$b^{*2}_{r+1} (z^*_{r+1} \cdot z^*_{r+1})$ $\vdots$ $b^{*2}_p (z^*_p \cdot z^*_p)$	1 $\vdots$ 1	$b^{*2}_{r+1} (z^*_{r+1} \cdot z^*_{r+1})$ $\vdots$ $b^{*2}_p (z^*_p \cdot z^*_p)$	$\sigma + \beta^{*2}_{r+1} (z^*_{r+1} \cdot z^*_{r+1})$ $\vdots$ $\sigma^2 + \beta^{*2}_p (z^*_p \cdot z^*_p)$	σ^2 $\vdots$ σ^2
Error	$S(x_\mu - \sum_{i=0}^{p} b_i z_{\mu i})^2$	N-p-1	$\dfrac{S(x_\mu - \sum_{i=0}^{p} b_i z_{\mu i})^2}{N-p-1}$	σ^2	σ^2
Total	$Sx_\mu^{\ 2}$	N			

Table 2A

Condensed Analysis of Variance Table in Terms of β^*-Weights

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Sum of Squares (MS)	Expected Value of MS	
				General	H_o
$\beta^*_o, \ldots, \beta^*_r$	$\sum_{i=0}^{r} b^*_i g^*_i$	$r+1$	$\sum_{i=0}^{r} \frac{b^*_i g^*_i}{r+1}$	$\sigma^2 + \sum_{i=0}^{r} \frac{\beta^{*2}_i (z^*_i \cdot z^*_i)}{r+1}$	$\sigma^2 + \sum_{i=0}^{r} \frac{\beta^{*2}_i (z^*_i \cdot z^*_i)}{r+1}$
$\beta^*_{r+1}, \ldots, \beta^*_p$	$\sum_{i=r+1}^{p} b^*_i g^*_i$	$p-r$	$\sum_{i=r+1}^{p} \frac{b^*_i g^*_i}{p-r}$	$\sigma^2 + \sum_{i=r+1}^{p} \frac{\beta^{*2}_i (z^*_i \cdot z^*_i)}{p-r}$	σ^2
Error	$S(x - \sum_{i=0}^{p} b_i z_{\mu i})^2$	$N-p-1$	$S(x_\mu - \sum_{i=0}^{p} b_i z_{\mu i})^2$	σ^2	σ^2
Total	Sx_μ^2	N			

To test H_o: $\beta^*_{r+1} = \ldots = \beta^*_p = 0$, compute

$$\frac{\sum_{r+1}^{p} b^*_i g^*_i \Big/ (p-r)}{S(x - \sum_{i=1}^{p} b_i z_i)^2} = \frac{\sum_{r+1}^{p} b^*_i g^*_i \Big/ (p-r)}{Sx_\mu^2 - \sum_{i=1}^{p} b_i g_i}$$

and compare with critical value of F.

From the Tables we see that the "sources" $\gamma_{r+1}, \ldots, \gamma_p$ are nonsignificant if and only if $\beta_{r+1}, \ldots, \beta_p$ are nonsignificant. This corresponds roughly to the statement that

$$\beta_{r+1} = \ldots = \beta_p = 0$$

if and only if

$$\gamma_{r+1} = \ldots = \gamma_p = 0 .$$

Theorem 3. The sum of sum of squares of error is

$$SSE = S(x_\mu - \hat{x}_\mu)^2 = \sum_{\mu=p+2}^{N} y_\mu^2$$

where $y = Ax$, as defined above.

Proof: From Theorem 5, Chapter III, we have

$$SSE = S(x_\mu - \hat{x}_\mu)^2 = Sx_\mu^2 - \sum_{i=0}^{p} b_i g_i$$

$$Sy_\mu^2 = y^T y = (Ax)^T(Ax) = x^T A^T x = x^T A^{-1} Ax = x^T x = Sx_\mu^2$$

$$\sum_{i=1}^{p+1} y_i^2 = \sum_{i=0}^{p} b*_i^2 (z*_i \cdot z*_i) = \sum_{i=0}^{p} b*_i g*_i = B*^T G* =$$

$$= (\alpha^{-1}B)^T(z*^T x) = B^T(\alpha^{-1})^T \alpha^T z^T x = B^T G = \sum_{i=0}^{b} b_i g_i$$

Therefore

$$SSE = \sum_{\mu=p+2}^{N} y_\mu^2$$

Theorem 4. The expected value of the sum of squares of error is $(N-p-1)\,\sigma^2$, and hence $\frac{SSE}{N-p-1}$ is unbiased estimate of σ^2.

Proof: From the preceding theorem the sum of squares of error is

$$SSE = \sum_{\mu=p+2}^{N} y_\mu{}^2$$

Hence

$$E(SSE) = E\left(\sum_{\mu=p+2}^{N} y_\mu{}^2\right) = \sum_{\mu=p+2}^{N} Ey_\mu{}^2 = (N-p-1)\,\sigma^2$$

The procedure for computing $y_i{}^2 = b^*_i{}^2(z^*_i \cdot z^*_i)$ will be discussed in the next chapter. First, however, we shall consider the problem of testing whether a linear combination of the β_i is equal to a hypothetical value, Λ_o.

Let $\Lambda = \beta^T \lambda$, $L = B^T \lambda$, where λ is a constant vector. Then

$$EL = E(B^T \lambda) = \beta^T \lambda = \Lambda$$

$$E(L-EL)^2 = E(B^T\lambda - \beta^T\lambda)^T(B^T\lambda - \beta^T\lambda) = \lambda^T E(B-\beta)^T(B-\beta) = \sigma^2 \lambda^T H^{-1}\lambda$$

L is a linear combination of the b_i, and hence of the y_i, which are normal. Therefore, L is normal (See Appendix). In fact

$$L = B^T\lambda = (\alpha B^*)^T\lambda = B^{*T}(\alpha^T\lambda) =$$

$$= \text{linear combination of } y_1, \ldots, y_{p+1}.$$

Thus $\dfrac{L-\Lambda}{\sigma(\lambda^T H^{-1}\lambda)^{\frac{1}{2}}}$ is a $N(0,1)$ variable, independent of $SSE = \sum_{\mu=p+2}^{N} y_\mu{}^2$

Hence, since $\frac{SSE}{\sigma^2}$ is chi-square with N-p-1 degrees of freedom, the statistic

$$\frac{\dfrac{L - \Lambda_o}{\sigma(\lambda^T H^{-1} \lambda)^{\frac{1}{2}}}}{\left[\dfrac{SSE}{\sigma^2 (N-p-1)}\right]^{\frac{1}{2}}} = \frac{\dfrac{L - \Lambda_o}{(\lambda^T H^{-1} \lambda)^{\frac{1}{2}}}}{\left[\dfrac{SSE}{N-p-1}\right]^{\frac{1}{2}}}$$

is a t-variable with N-p-1 degrees of freedom under the hypothesis that $\Lambda = \Lambda_o$. It may be compared with critical values from the tables.

In the special case where $\Lambda = \beta_i$ all $\lambda_j = 0$ except $\lambda_j = 1$. In this case the statistic above becomes

$$\frac{\dfrac{b_i - \beta_{i0}}{(c_{ii})^{\frac{1}{2}}}}{\left(\dfrac{SSE}{N-p-1}\right)^{\frac{1}{2}}}$$

where $H^{-1} = C = (c_{ij})$, which under the hypothesis that $\beta_i = \beta_{i0}$ is distributed by the t distribution with N-p-1 degrees of freedom.

CHAPTER V

COMPUTATIONAL PROCEDURES

In this chapter we shall develop computational procedures (based essentially on the elimination method of Gauss) which render the application of the estimation and testing methods of the last two chapters routine.

It is easiest to begin the algorithm using the data for the preliminarily orthogonalized model, rather than the original model. Thus we have $x = z' \beta' + \varepsilon$ where $z' = [z_o, z'_1, \ldots, z'_p]$. The normal equations for this model are

$$H'B' = G' ,$$

where

$$H = \left[\begin{array}{c:c} N & 0 \\ \hdashline 0 & h \end{array}\right], \quad G' = \left[\begin{array}{c} N\bar{x} \\ \hdashline g \end{array}\right]. \quad B' = \left[\begin{array}{c} b'_o \\ \hdashline n \end{array}\right]; \; h_{ij} = (z'_i \ldots z'_j)$$

or we may write the normal equation as

$$hb = g, \quad b'_o = \bar{x} .$$

Since the solutions to HB G for the original model are

$$B = \begin{bmatrix} b_o \\ b_1 \\ \vdots \\ b_p \end{bmatrix} = \left[\begin{array}{c} b'_o - \sum_{i=1}^{p} b_i \bar{z}_o \\ \hdashline b \end{array}\right]$$

it is clear that no difficulty is involved in recovering solutions for the original model.

So we begin by forming the matrices

$$\overset{p\times p}{h} = (h_{ij}) \quad \text{and} \quad \overset{p\times 1}{g} = (g_i)$$

where

$$h_{ij} = (z'_i \cdot z'_j) = S(z_{\mu i} - \bar{z}_i)(z_{\mu j} - \bar{z}_j)$$

$$S\, z_{\mu i}\, z_{\mu j} - \frac{Sz_{\mu i} Sz_{\mu j}}{N}$$

$$g_i = (z'_i \cdot x) = S(z_{\mu i} - \bar{z}_i)\, x_\mu = Sz_{\mu i} x_\mu - \frac{Sz_{\mu i} Sx_\mu}{N}$$

Also, $(z'_i \cdot x') = S(z_{\mu i} - \bar{z}_i)\,(x_\mu - \bar{x}) = Sz_{\mu i} x_\mu - \dfrac{Sz_{\mu i} Sx_\mu}{N}$

Thus an alternate definition of g is $g = z'^T x'$.

The whole of the procedure is contained in the elimination method for solving the system of equations

$$hb = g = Ig$$

where we also construct the inverse of h by performing on I those same transformations which reduce h to I. This can be done since the linear independence of the z_i implies that h is nonsingular. The things we wish to exhibit in the sequence of tableaux, which form the reduction of h to I, are listed below:

(a) The regression estimates for a sequence of p fittings where the kth fitting requires the least squares estimation of $\beta_1, \ldots, \beta_k$ using only the first k + 1 vectors $z_0, z'_1, \ldots, z'_p$;

(b) The corresponding inverse matrices;

(c) The transformation α^* which connects the model

$$x = z' \ \beta' + \varepsilon$$

to the completely orthogonal model

$$x = z^* \ \beta^* + \varepsilon$$

where $z^* = z' \alpha^*$, $\beta^* = (\alpha^*)^{-1}\beta$, z^* being previously defined;

(d) The inner products $(z^*_i \cdot z^*_i)$ which appear in the analysis of variance tables in Chapter IV; and

(e) The estimates b^*_i $(i = 1, \ldots, p)$, which also occur in the analysis of variance tables.

We shall need the two following theorems:

Theorem 1. The matrix h is positive definite.

Proof: Let $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}$ be an arbitrary vector. Then

$$\alpha^T H = \alpha^T z^T z \alpha = S\left(\sum_{i=1}^{p} z_{\mu 1} \alpha_i\right)^2 \geq 0$$

Equality can hold if and only if $\sum_{i=1}^{p} z_i \alpha_i = 0$. However, since the z_i are linearly independent, $\sum_{i=1}^{p} z_i \alpha_i = 0$ if and only if $\alpha = 0$. Thus $\alpha^T H\alpha \geq 0$, and $\alpha^T H\alpha = 0$ if and only if $\alpha = 0$. Hence H is positive definite.

Definition 1. The Gundelfinger determinants of the matrix h are the determinants of the submatrices

$$h_k = \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & & \vdots \\ h_{k1} & \cdots & h_{kk} \end{bmatrix}, \quad k = 1, 2, \ldots, p. \qquad (5.1)$$

<u>Theorem 2.</u> The Gundelfinger determinants of h are greater than zero.

Proof: We apply the Gram-Schmidt process described in Chapter II to the vectors $z'_1, \ldots, z'_k$. These vectors are related to the orthogonal vectors $z^*_1, \ldots, z^*_k$ by

$$\zeta^*_k = \zeta'_k \; \alpha_k ,$$

where

$$\zeta^*_k = \left[z^*_1, \ldots, z^*_k \right], \quad \zeta'_k = \left[z'_1, \ldots, z'_k \right]$$

$$\alpha_k = \overset{k \times k}{\begin{bmatrix} 1 & & \text{———} \\ 0 & \ddots & \text{——} \\ | & \ddots & \vdots \\ 0 & \text{——} \; 0 & 1 \end{bmatrix}} ,$$

α_k thus having a determinant equal to one.

Now

$$h_k = \zeta'^T_k \, \zeta'_k = (\zeta^* \alpha_k^{-1})^T \, (\zeta^* \alpha_k^{-1}) = \alpha_k \, \zeta^{*T} \zeta^* \alpha_k^{-1}$$

$$= \alpha_k \, h^*_k \; \alpha_k^{-1} \quad ,$$

where h^*_k is a diagonal matrix with $(z^*_i \cdot z^*_i)$ in the ith diagonal position $(i = 1, 2, \ldots, k)$. Hence

$$| h_k | = |\alpha_k| \; |h^*_k| \; |\alpha_k^{-1}| = | h^*_k |$$

That is

$$| h_k | = | h^*_k | = \begin{vmatrix} (z^*_1 \cdot z^*_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & (z^*_k \cdot z^*_k) \end{vmatrix} = (z^*_1 \cdot z^*_1) \ldots (z^*_k \cdot z^*_k)$$

But according to Theorem 1, Chapter II, $(z^*_i \cdot z^*_i) > 0$ for each $i = 1,\ldots,p$. Thus $|h_k| > 0$.

The calculations which reduce h to I are schematized in the following sequence of tableaux, which, as we shall prove, yield all the information required above. The starting, or zero, tableau is the p x (2p + 1) matrix

$$T^{(o)} = [h,g,I]$$

For $k = 1,\ldots,p$ $T^{(k)}$ is obtained from $T^{(k-1)}$ by performing on $T^{(k-1)}$ the elementary row transformations which reduce the kth column to a unit vector with unity in the kth position. As we shall show below, these transformations will not require the interchange of rows. The final result will be

$$T^{(p)} = [I,b,c]$$

where $c = h^{-1}$.

Before setting down a description of the schematization, we shall prove that the above rule does not fail. The only way it can fail at some stage, say the $(k-1)^{\underline{st}}$ where k might be 1,2, or p, is for the pivot element $h_{kk}^{(k-1)}$ to be zero. In that case it would be impossible to divide by that element to obtain unity in that position.

<u>Theorem 3</u>. $h_{kk}^{(k-1)}$ is not equal to zero.

Proof: Suppose $h_{kk}^{(k-1)} = 0$. In that case, we would have, as the k x k submatrix in the upper left-hand corner of $T^{(k-1)}$

$$h_k^{(k-1)} = \begin{bmatrix} 1 & \ldots & 0 & h_{1k}^{(k-1)} \\ \vdots & & \vdots & \vdots \\ 0 & \ldots & 1 & h_{k-1,k}^{(k-1)} \\ 0 & \ldots & 0 & 0 \end{bmatrix}$$

where $h_k^{(k-1)}$ is obtained by elementary row transformations from

$$h_k = \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & & \vdots \\ h_{k1} & \cdots & h_{kk} \end{bmatrix}$$

Obviously the determinant of $h_k^{(k-1)}$ vanishes. But

$$h_k^{(k-1)} = Ph_k$$

where P is nonsingular. Also, h_k is nonsingular since $|h_k|$ is a Gundelfinger determinant. Since the product of nonsingular matrices is nonsingular, $h_k^{(k-1)}$ is nonsingular. Thus the supposition that $h_{kk}^{(k-1)} = 0$ leads to a contradiction. Hence for every k, $k = 1,2,\ldots,p$, $h_{kk}^{(k-1)} \neq 0$.

The Gaussian tableaux are exhibited in Table 3. For the kth fitting

$$I_k = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

is the identity matrix to which h_k is reduced.

Corresponding to the full regression model

$$x = z' \quad \beta' + \varepsilon$$

where all points of the structural vectors z_o, $z_1,,,.z'_p$ are used, we define for $k = 1,2,\ldots,$ p-1

$$x = z'(k) \cdot \beta'(k) + \varepsilon(k) \qquad (5.2)$$

where $z(k) = [z_o z'_1 \ldots z'_k \ldots z'_p]$, $\beta'(k) = \begin{bmatrix} \beta'_o \\ \vdots \\ \beta'_k \end{bmatrix}$.

Corresponding to each of these models, $k = 1,2,\ldots,p-1$, are normal equations,

Table 3

The Gaussian Tableaux

Fitting No.	Equation No.	Successive Reductions of h	Successive fittings	Successive inverses
Initial data $T^{(o)}$	$(0)_1$ $\vdots$ $(0)_p$	$\begin{matrix} h_{11} & \cdots & h_{1p} \\ \vdots & & \vdots \\ h_{p1} & \cdots & h_{pp} \end{matrix}$	$\begin{matrix} g_1 \\ \vdots \\ g_p \end{matrix}$	$\begin{matrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{matrix}$
First fitting $T^{(1)}$	$(1)_1$ $(1)_2$ $\vdots$ $(1)_p$	$\begin{matrix} [1] & [h_{12}^{(1)}] & \cdots & h_{1p}^{(2)} \\ 0 & h_{22}^{(1)} & \cdots & h_{2p}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & h_{p2}^{(1)} & \cdots & h_{pp}^{(2)} \end{matrix}$	$\begin{matrix} [b_1^{(1)}] \\ g_2^{(1)} \\ \vdots \\ g_p^{(1)} \end{matrix}$	$\begin{matrix} [c_{11}^{(1)}] & 0 & \cdots & 0 \\ c_{21}^{(1)} & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ c_{p1}^{(1)} & 0 & \cdots & 1 \end{matrix}$
Second fitting $T^{(2)}$	$(2)_1$ $(2)_2$ $(2)_3$ $\vdots$ $(2)_p$	$\begin{matrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} h_{13}^{(2)} \\ h_{23}^{(2)} \end{bmatrix} & \begin{matrix} \cdots & h_{1p}^{(2)} \\ \cdots & h_{2p}^{(2)} \end{matrix} \\ \begin{matrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{matrix} & \begin{matrix} h_{33}^{(2)} \\ \vdots \\ h_{p3}^{(2)} \end{matrix} & \begin{matrix} \cdots & h_{3p}^{(2)} \\ & \vdots \\ \cdots & h_{pp}^{(2)} \end{matrix} \end{matrix}$	$\begin{matrix} \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \end{bmatrix} \\ g_3^{(2)} \\ \vdots \\ g_p^{(2)} \end{matrix}$	$\begin{matrix} \begin{bmatrix} c_{11}^{(2)} & c_{12}^{(2)} \\ c_{21}^{(2)} & c_{22}^{(2)} \end{bmatrix} & \begin{matrix} 0 & \cdots & 0 \\ 1 & \cdots & 0 \end{matrix} \\ \begin{matrix} c_{31}^{(2)} & c_{32}^{(2)} \\ \vdots & \vdots \\ c_{p1}^{(2)} & c_{p2}^{(2)} \end{matrix} & \begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{matrix} \end{matrix}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(Continued)

Table 3 (Continued)

Fitting No.	Equation No.	Successive Reductions of h	Successive fittings	Successive inverses
(k-1)st fitting $T^{(k-1)}$	$\begin{matrix}(k-1)_1\\ \vdots\\ (k-1)_{k-1}\\ (k-1)_k\\ \vdots\\ (k-1)_p\end{matrix}$	$\begin{array}{cccc} \begin{bmatrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&1\end{bmatrix} & \begin{bmatrix}h_{1k}^{(k-1)}\\ \vdots\\ h_{k-1,k}^{(k-1)}\end{bmatrix} & \cdots & \begin{matrix}h_{1p}^{(k)}\\ \vdots\\ h_{k-1,p}^{(k)}\end{matrix} \\ \begin{matrix}0&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&0\end{matrix} & \begin{matrix}h_{kk}^{(k-1)}\\ \vdots\\ h_{pk}\end{matrix} & \cdots & \begin{matrix}h_{kp}^{(k)}\\ \vdots\\ h_{pp}^{(k)}\end{matrix}\end{array}$	$\begin{matrix}\begin{bmatrix}b_1^{(k-1)}\\ \vdots\\ b_{k-1}^{(k-1)}\end{bmatrix}\\ g_k^{(k-1)}\\ \vdots\\ g_p^{(k-1)}\end{matrix}$	$\begin{array}{cc} \begin{bmatrix}c_{11}^{(k-1)}&\cdots&c_{1(k-1)}^{(k-1)}\\ \vdots&&\vdots\\ c_{k-1,1}^{(k-1)}&\cdots&c_{k-1,k-1}^{(k-1)}\end{bmatrix} & \begin{matrix}0&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&0\end{matrix} \\ \begin{matrix}c_{k1}^{(k-1)}&\cdots&c_{k,k-1}^{(k-1)}\\ \vdots&&\vdots\\ c_{p1}^{(k-1)}&\cdots&c_{p,k-1}^{k-1}\end{matrix} & \begin{matrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&1\end{matrix}\end{array}$
k^{th} fitting $T^{(k)}$	$\begin{matrix}(k)_1\\ \vdots\\ (k)_k\\ (k)_{k+1}\\ \vdots\\ (k)_p\end{matrix}$	$\begin{array}{cccc} \begin{bmatrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{bmatrix} & \begin{bmatrix}h_{1,k+1}^{(k)}\\ \vdots\\ h_{k,k+1}^{(k)}\end{bmatrix} & \cdots & \begin{matrix}h_{1p}^{(k)}\\ \vdots\\ h_{kp}^{(k)}\end{matrix} \\ \begin{matrix}0&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&0\end{matrix} & \begin{matrix}h_{k+1,k+1}^{(k)}\\ \\ h_{p,k+1}\end{matrix} & \cdots & \begin{matrix}h_{k+1,p}\\ \vdots\\ h_{kp}\end{matrix}\end{array}$	$\begin{matrix}\begin{bmatrix}b_1^{(k)}\\ \\ b_k^{(k)}\end{bmatrix}\\ g_{k+1}^{(k)}\\ \vdots\\ g_p^{(k)}\end{matrix}$	$\begin{array}{cc} \begin{bmatrix}c_{11}^{(k)}&\cdots&c_{1k}^{(k)}\\ \vdots&&\vdots\\ c_{k1}^{(k)}&\cdots&c_{kk}^{(k)}\end{bmatrix} & \begin{matrix}0&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&0\end{matrix} \\ \begin{matrix}c_{k+1,1}^{(k)}&\cdots&c_{k+1,k}^{(k)}\\ \vdots&&\vdots\\ c_{p1}^{(k)}&\cdots&c_{pk}\end{matrix} & \begin{matrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{matrix}\end{array}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
pth fitting $T^{(p)}$	$\begin{matrix}(p)_1\\ \vdots\\ (p)_p\end{matrix}$	$\begin{matrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&1\end{matrix}$	$\begin{matrix}b_1^{(p)}\\ \vdots\\ b_p^{(p)}\end{matrix}$	$\begin{matrix}c_{11}^{(p)}&\cdots&c_{1p}^{(p)}\\ \vdots&&\vdots\\ c_{p1}^{(p)}&\cdots&c_{pp}^{(p)}\end{matrix}$

$$h_k\, b(k) = g(k), \quad b'_o = \bar{Y}\ , \ k = 1,2,\ldots p - 1 \tag{5.3}$$

where h_k was defined in (5.1) as the k x k matrix in the upper lefthand corner of h, and $g(k)$ consists of the first k components of the vector g.

Consider again the kth tableau

$$\left.\begin{array}{cccc|c|ccc}
\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} & \begin{matrix} h^{(k)}_{1,k\,1} \\ \\ h_{k1,k+1} \end{matrix} & \begin{matrix} \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} h^{(k)}_{1p} \\ \vdots \\ h^{(k)}_{k+p} \end{matrix} & \begin{bmatrix} b^{(k)}_1 \\ \vdots \\ b^{(k)}_k \end{bmatrix} & \begin{bmatrix} c^{(k)}_{11} & \cdots & c^{(k)}_{1k} \\ \vdots & & \vdots \\ c^{(k)}_{k1} & \cdots & c^{(k)}_{kk} \end{bmatrix} & & \begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix} \\
\begin{matrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{matrix} & \begin{matrix} h^{(k)}_{k+1,k+1} \\ \vdots \\ h^{(k)}_{p,k+1} \end{matrix} & \begin{matrix} \cdots \\ \\ \cdots \end{matrix} & \begin{matrix} h^{(k)}_{k+1,p} \\ \vdots \\ h^{(k)}_{pp} \end{matrix} & \begin{matrix} g^{(k)}_{k\,1} \\ \vdots \\ g^{(k)}_p \end{matrix} & \begin{matrix} c^{(k)}_{k+1,1} & \cdots & c_{k+1,pk} \\ \vdots & & \vdots \\ c^{(k)}_{p1} & \cdots & c^{(k)}_{p1} \end{matrix} & & \begin{matrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{matrix}
\end{array}\right.$$

Clearly, the solution to (5.3) is

$$b(k) = \begin{bmatrix} b_1^{(k)} \\ \vdots \\ b_k^{(k)} \end{bmatrix}$$

Since the elementary transformations which reduced h_k to I_k transformed I_k to

$$c_k = \begin{bmatrix} c_{11}^{(k)} & \cdots & c_{1k}^{(k)} \\ \vdots & & \vdots \\ c_{k1}^{(k)} & \cdots & c_{kk}^{(k)} \end{bmatrix}$$

c_k is the inverse of h_k. Thus the successive tableaux display (a) and (b).

Next, let us prove the following lemma.

Lemma. For $i = 1,2,\ldots, k-1$,

$$\sum_{j=1}^{k-1} h_{ij}\ h_{jk}^{(k-1)} = h_{ik}$$

Proof: Let

$$H = \begin{bmatrix} h_{11} & \cdots & h_{1,k-1} \\ \vdots & & \vdots \\ h_{k-1,1} & \cdots & h_{k,k-1} \end{bmatrix} \quad H_k^{(k-1)} = \begin{bmatrix} h_{1k}^{(k-1)} \\ \vdots \\ h_{k-1,k}^{(k-1)} \end{bmatrix} \quad H_k = \begin{bmatrix} h_{1k} \\ \vdots \\ h_{k-1,k} \end{bmatrix} \quad (5.4)$$

From the Gauss procedure, we have

$$\begin{bmatrix} H^{-1} & 0 \\ x & x \end{bmatrix} \begin{bmatrix} H & H_k \\ H_k^T & h_{kk} \end{bmatrix} = \begin{bmatrix} I_{k-1} & H_k^{(k-1)} \\ 0 & h_{kk}^{(k-1)} \end{bmatrix}$$

Hence

$$H^{-1} H_k = H_k^{(k-1)}$$

or

$$H_k = HH_k^{(k-1)}$$

Thus for $i = 1, 2, \ldots, k-1$

$$\sum_{j=1}^{k-1} h_{ij} h_{jk}^{(k-1)} = h_{ik}$$

Relative to (c), we shall now prove the following theorem.

Theorem 4. The orthogonal vectors $z^*_0, \ldots z^*_p$ formed by the Gram-Schmidt process may be written down from the successive tableaux. In fact

$$z^*_k = z'_k - \sum_{i=1}^{k-1} h_{ik}^{(k-1)} z'_i$$

Proof: Let

$$\zeta^*_k = z'_k - \sum_{i=1}^{k-1} h_{ik}^{(k-1)} z'_i$$

Then, for $m \neq k$, say $m < k$

$$(\zeta^*_k \cdot \zeta^*_m) = \left(\left[z'_k - \sum_{i=1}^{k-1} h_{ik}^{(k-1)} z'_i\right] \cdot \left[z'_m - \sum_{j=1}^{m-1} h_{jm}^{(m-1)} z'_j\right]\right) =$$

$$= (z'_k \cdot z'_m) - \sum_{i=1}^{k-1} h_{ik}^{(k-1)} (z'_i \cdot z'_m) - \sum_{j=1}^{m-1} h_{jm}^{(m-1)} (z'_k \cdot z'_j)$$

$$+ \sum_{j=1}^{m-1} h_{jm}^{(m-1)} \left[\sum_{i=1}^{k-1} h_{ik}^{(k-1)} (z'_i \cdot z'_j)\right]$$

Applying the lemma

$$(\zeta^*_k \cdot \zeta^*_m) = h_{km} - h_{km} - \sum_{j=1}^{m-1} h_{jm}^{(m-1)} h_{kj} + \sum_{j=1}^{m-1} h_{jm}^{(m-1)} h_{kj} = 0$$

Also, for m still less than k,

$$(\zeta^*_k \cdot z'_m) = (z'_k \cdot z'_m) - \sum_{i=1}^{k-1} h_{ik}^{(-1)} (z'_i \cdot z'_m) = h_{km} - h_{km} = 0$$

But

$$(z^*_k \cdot z'_m) = \left(z^*_k \cdot \left[z^*_m + \sum_{i=1}^{m-1} \frac{(z^*_i \cdot z'_m)}{(z^*_i \cdot z^*_i)} z^*_i\right]\right) =$$

$$= (z^*_k \cdot z^*_m) + \sum_{i=1}^{m-1} \frac{(z^*_i \cdot z'_m)}{(z^*_i \cdot z^*_i)} (z^*_k \cdot z^*_i) = 0$$

since $(z^*_i \cdot z^*_j) = 0$, for $i \neq j$. Thus to show that $z^* = \zeta^*$ we consider the two systems of equations

$$0 = (z'_i \cdot z^*_k) = \left(z'_i \cdot \left[z'_k - \sum_{j=1}^{k-1} a_{jk} z'_j\right]\right)$$

$$0 = (z'_i \cdot \zeta^*_k) = \left(z'_i \cdot \left[z'_k - \sum_{j=1}^{k-1} h_{jk}^{(k-1)} \, z'_j \right] \right)$$

for $i = 1,2,\ldots, k-1$

These simplify to

$$\sum_{j=1}^{k-1} h_{ij} \, a_{jk} = h_{ik} \qquad i = 1,2,\ldots, k-1$$

$$\sum_{j=1}^{k-1} h_{ij} \, h_{jk}^{(k-1)} = h_{ik} \quad i = 1,2,\ldots, k-1$$

The two systems of equations have the same unique solution, i.e.,

$$a_{jk} = h_{jk}^{(k-1)}, \quad j = 1,2,\ldots, k-1$$

Therefore

$$z^*_k = \zeta^*_k = z'_k - \sum_{j=1}^{k-1} h_{jk}^{(k-1)} \, z'_j$$

We now resolve question (d)

<u>Lemma</u>. If A is a positive definite k x k matrix with inverse c and if A^*_{ij} is the matrix obtained from A by deleting the ith row and jth column, then

$$\frac{1}{c_{kk}} = a_{kk} - [a_{k1} \cdots a_{k,k-1}] \, (A^*_{kk})^{-1} \begin{bmatrix} a_{1k} \\ \vdots \\ a_{k-1,k} \end{bmatrix}$$

Proof: The inverse element c_{kk} may be written in the form

$$c_{kk} = \frac{|A^*_{kk}|}{|A|}$$

$|A|$ may be expanded by minors

$$|A| = \sum_{i=1}^{k} (-1)^{i+k} a_{ki} \, |A^*_{ki}|$$

Hence

$$\frac{1}{c_{kk}} = a_{kk} + \sum_{i=1}^{k-1} (-1)^{i+k} a_{ki} \frac{|A^*_{ki}|}{|A^*_{kk}|}$$

However

$$A^*_{ki} = \sum_{j=1}^{k-1} (-1)^{j+k-1} a_{jk} \, |A^*_{ki \cdot jk}| = \sum_{j=1}^{k-1} (-1)^{j+k-1} a_{jk} |A^*_{kk \cdot ji}|$$

Thus

$$\frac{1}{c_{kk}} = a_{kk} - \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} (-1)^{i+j} a_{ki} \, a_{jk} \, \frac{|A^*_{kk \cdot ji}|}{|A^*_{kk}|}$$

However, $(-1)^{i+j} \dfrac{|A^*_{kk \cdot ji}|}{|A^*_{kk}|}$ is an inverse element of A^*_{kk}. Thus

$$\frac{1}{c_{kk}} = a_{kk} - [a_{k1} \cdots a_{k,k-1}] \, (A^*_{kk})^{-1} \begin{bmatrix} a_{k1} \\ \vdots \\ a_{k,k-1} \end{bmatrix}$$

<u>Theorem 5</u>. $(z^*_k \cdot z^*_k) = \dfrac{1}{c^{(k)}_{kk}}$

Proof: By Theorem 4

$$z^*_k = [z'_1 \cdots z'_k] \begin{bmatrix} -h^{(k-1)}_{1k} \\ \vdots \\ -h^{(k-1)}_{k-1,k} \\ 1 \end{bmatrix}$$

Hence

$$(z^*_k \cdot z^*_k) = \begin{bmatrix} -h_{1k}^{(k-1)} & \cdots & -h_{k-1,k}^{(k-1)} & 1 \end{bmatrix} \begin{bmatrix} h_{11} & \cdots & h_{1k} \\ \vdots & & \vdots \\ h_{k1} & \cdots & h_{kk} \end{bmatrix} \begin{bmatrix} -h_{1k}^{(k-1)} \\ \vdots \\ -h_{k-1,k} \\ 1 \end{bmatrix}$$

or, using the notation in (5.4)

$$(z^*_k \cdot z^*_k) = \begin{bmatrix} H_k^{(k-1)\ T}, & 1 \end{bmatrix} \left[\begin{array}{c|c} H & H_k \\ \hline H^T & h_{kk} \end{array}\right] \left[\begin{array}{c} -H_k^{(k-1)} \\ \hline 1 \end{array}\right]$$

$$= \left[H_k^{(k-1)}\right]^T H \ \ H_k^{(k-1)} - H_k^T H_k^{(k-1)} - \left[H_k^{(k-1)}\right]^T H_k + h_{kk}$$

$$= h_{kk} - 2H_k^T H_k^{(k-1)} + \left[H_k^{(k-1)}\right]^T H \ \ H_k^{(k-1)}$$

However, by the lemma to Theorem 4,

$$H_k^T = \left[H_k^{(k-1)}\right]^T H \ .$$

Hence

$$(z^*_k \cdot z^*_k) = h_{kk} - H_k^T H^{-1} H_k$$

We now apply the lemma to the matrix h_k to obtain

$$\frac{1}{c_{kk}^{(k)}} = h_{kk} - H_k^T H^{-1} H_k$$

It is thus clear that

$$(z^*_k \cdot z^*_k) = \frac{1}{c_{kk}^{(k)}}$$

Relative to (e) we have

<u>Theorem 6</u>. Let

$$x = z^*(k)\ B^*(k) + \varepsilon(k)$$

be the completely orthogonal model corresponding to the first k + 1 vectors $z_o, z_1, \ldots, z_k$. Then

$$B^*(k) = \begin{bmatrix} (z^*_o \cdot x)/(z^*_o \cdot z^*_o) \\ \vdots \\ (z^*_k \cdot x)/(z^*_k \cdot z^*_k) \end{bmatrix} = \begin{bmatrix} b^*_o \\ \vdots \\ b^*_k \end{bmatrix}$$

and furthermore $b^*_k = b_k^{(k)}$.

The theorem thus states that the first k + 1 estimates $b^*_o, \ldots, b^*_k$ obtained for the orthogonal equivalent of the pth fitting are the same as those obtained for the orthogonal equivalent of the lower order (kth) fitting. Also the theorem shows us where it is in the succession of tableaux that we find the b^*_i, which are required in the analysis of variance tables of Chapter IV.

Proof: From analogy with the development in Chapter II, we see that the least squares estimates of $\beta^*_o, \beta^*_1, \ldots, \beta^*_k$ are

$$b^*_i(k) = \frac{(z^*_i(k) \cdot x)}{(z^*_i(k) \cdot z^*_i(k))}$$

However, since z^*_i depends only on $z_o, \ldots z_i$, for $i \leq k$

$$z^*_i(k) = z'_i - \sum_{j=1}^{i-1} b_{ji}^{(i-1)} z'_j = z^*_i \ .$$

Hence $b^*_i(k) = \dfrac{(z^*_i \cdot x)}{(z^*_i \cdot z^*_i)} = b^*_i$

We have, for $k = 1, 2, \ldots, p-1$.

$$z^*(k) = \alpha^*(k)\ z'(k)$$

and hence

$$\alpha^*(k)\ B^*(k) = B'(k)$$

or

$$\begin{bmatrix} 1 & \alpha^*_{00} & \cdots & \alpha^*_{0k} \\ 0 & 1 & \cdots & 1k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} b^*_0 \\ b^*_1 \\ \vdots \\ b^*_k \end{bmatrix} \begin{bmatrix} b_0^{(k)} \\ b_1^{(k)} \\ \vdots \\ b_k^{(k)} \end{bmatrix}$$

From this we see that $b^*_k = b_k^{(k)}$, for $k = 1, 2, \ldots, p-1$. By the same reasoning $b^*_p = b_p$. We now wish to achieve a final simplification of the analysis of variance table, Table 2A, when no need exists for breaking down the sum of squares, Sx^2, into individual independent components, $y_i^2 = b^*_i g^*_i$.

Consider the r^{th} fitting when the first $r + 1$ vectors $z_0, \ldots, z_r$ are employed in the model:

$$x = z(r)\,\beta(r) + \varepsilon(r)$$

Its set of normal equations we write as $H_r B(r) = G(r)$. Its orthogonal equivalent we write

$$x = z^*(r)\ \beta^*(r) + \varepsilon(r) \quad ,$$

with the corresponding normal equations

$$b^*_i(r) = \frac{(z^*_i(r)\cdot x)}{(z^*_i(r)\cdot z^*_i(r))} \quad , \quad i = 0, 1, \ldots r$$

As a consequence of the above theorem, we see that

$$b^*_i(r) g^*_i(r) = b^*_i g^*_i$$

Also, as a corollary to Theorem 3, Chapter II, the sum of these products is invariant for equivalent models. Write

$$(SSR)_r = \sum_{i=0}^{r} b_i(r)\ g_i(r)$$

the reduction in sum of squares effected by a fitting of order r. We thus have

$$\sum_{i=0}^{r} b^*_i g^*_i = (SSR)_r ,$$

which can be obtained directly from the solution to the rth fitting. Similarly, defining

$$(SSR)_p = \sum_{i=0}^{p} b_i g_i = \sum_{i=0}^{p} b^*_i g^*_i$$

as the reduction effected by the pth fitting, we obtain by subtraction,

<u>Corollary</u>. $\sum_{i=r+1}^{p} b^*_i g^*_i = (SSR)_p - (SSR)_r$. Thus Table 2A can be written in the form of Table 2B.

The F statistic for testing whether $\beta_{r+1} = \ldots = \beta_p = 0$ becomes

$$\frac{[(SSR)_p - (SSR)_r]/(p-r)}{[Sx^2 - (SSR)_p]/(N-p-1)}$$

which under H_o has the F distribution with (p-r) and (N-p-1) degrees of freedom.

Table 2B

Condensed Analysis of Variance Table in Terms of Original Weights

Source	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Sum of Squares (MS)	Expected Value of MS	
				General Hypothesis	H_o
Mean	$y_1^2 = N\bar{x}^2$	1	$N\bar{x}^2$	$\sigma^2 + \frac{(\beta^*_o)^2}{N}$	$\sigma^2 + \frac{(\beta^*_o)^2}{N}$
$\beta^*_1 \cdots \beta^*_r$	$(SSR)_r - N\bar{x}^2 = \sum_{i=2}^{r+1} y_i^2$	r	$\frac{(SSR)_r - N\bar{x}^2}{r}$	$\sigma^2 + \sum_{i=1}^{r+2} \beta^{*2}_i (z^*_i \cdot z^*_i)$	$\sigma^2 + \sum_{i=1}^{r+1} \beta^{*2}_i (z^*_i \cdot z^*_i)$
$\beta^*_{r+1} \cdots \beta^*_p$	$(SSR)_p - (SSR)_r = \sum_{i=r+2}^{p+1} y_i^2$	p-r	$\frac{(SSR)_p - (SSR)_r}{p-r}$	$\sigma^2 + \sum_{i=r+1}^{p} \beta^{*2}_i (z^*_i \cdot z^*_i)$	σ^2
Error	$Sx^2 - (SSR)_r = \sum_{i=p+2}^{N} y_i^2$	N-p-1	$\frac{Sx^2 - (SSR)_p}{N-p-1}$	σ^2	σ^2

Now let us illustrate the theory by a numerical example. Woltz, Reid, and Colwell (9) analyzed the regression of nicotine on certain minerals in tobacco leaf. Some of their data are given in Table 4. This problem was also considered by Anderson and Bancroft (1).

From the table the total sum of squares is

$$Sx^2 = 103.9112$$

The means are

$$\bar{x} = 1.9968$$
$$\bar{z}_1 = 2.2752$$
$$\bar{z}_2 = .7464$$
$$\bar{z}_3 = 3.4800$$
$$\bar{z}_4 = 2.2936$$

The initial data for the Gaussian elimination procedure are

$$h_{11} = (z'_1 \cdot z'_1) = 1.6500$$
$$h_{12} = (z'_1 \cdot z'_2) = 0.3349$$
$$h_{13} = (z'_1 \cdot z'_3) = 1.4496$$
$$h_{14} = (z'_1 \cdot z'_4) = 1.4496$$
$$h_{22} = (z'_2 \cdot z'_2) = 0.2054$$
$$h_{23} = (z'_2 \cdot z'_3) = 0.1270$$
$$h_{24} = (z'_2 \cdot z'_4) = -0.2287$$
$$h_{33} = (z'_3 \cdot z'_3) = 5.6306$$
$$h_{34} = (z'_3 \cdot z'_4) = -3.0010$$
$$h_{44} = (z'_4 \cdot z'_4) = 7.4842$$

Table 4

Percentage of Certain Materials Found

in Tobacco Leaf

x	z_1	z_2	z_3	z_4
1.50	1.86	0.93	3.22	1.42
2.65	2.77	0.74	4.68	0.97
1.80	2.04	0.82	3.24	1.53
1.80	2.16	0.75	3.37	2.32
2.48	2.50	0.63	3.72	2.04
1.46	2.07	0.63	2.74	2.38
2.47	2.74	0.62	3.54	2.70
1.75	2.15	0.76	3.24	2.24
1.52	2.06	0.88	3.48	1.90
2.72	2.72	0.66	4.33	1.70
1.33	2.16	0.87	3.00	2.74
1.72	2.35	0.72	3.06	1.80
2.72	2.16	0.62	3.55	2.25
1.58	1.82	0.80	2.78	3.00
2.66	2.74	0.60	3.46	3.43
2.12	2.32	0.79	3.07	2.76
2.11	2.38	0.76	2.96	2.92
1.79	2.32	0.69	3.57	2.76
2.16	2.20	0.80	3.30	2.54
1.94	2.22	0.71	3.41	2.73
1.70	2.06	0.82	4.42	2.15
1.91	2.08	0.74	3.97	2.22
1.87	2.33	0.90	3.77	2.68
1.94	2.23	0.72	3.39	2.25
2.22	2.44	0.70	3.73	1.91

x Percentage of Nicotine

z_1 Percentage of Nitrogen

z_2 Percentage of Phosphorus

z_3 Percentage of Calcium

z_4 Percentage of Potassium

The tableaux are given in the table below.

Table 5

The Gaussian Tableaux for the

Tobacco Leaf Problem

	Successive Reductions of H				Successive Fittings	Successive Inverses			
	h				g	I			
$(0)_1$	1.650	-0.335	1.450	0.019	2.075	1	0	0	0
$(0)_2$	-0.335	0.205	-0.127	-0.229	-0.600	0	1	0	0
$(0)_3$	1.450	-0.127	5.631	-3.001	2.512	0	0	1	0
$(0)_4$	0.019	0.229	-3.001	7.484	-0.260	0	0	0	1
$(1)_1$	[1]	-0.203	0.879	0.012	[1.258]	[0.606]	0	0	0
$(1)_2$	0	0.137	0.167	-0.225	-0.179	0.203	1	0	0
$(1)_3$	0	0.167	4.357	-3.018	0.689	-0.879	0	1	0
$(1)_4$	0	-0.225	-3.018	7.484	-0.284	-0.012	0	0	0
$(2)_1$	1	0	1.256	-0.320	0.994	0.906	1.477	0	0
$(2)_2$	0	1	1.217	-1.636	-1.301	1.477	7.278	0	0
$(2)_3$	0	0	4.154	-2.744	0.906	-1.126	-1.218	1	0
$(2)_4$	0	0	-2.744	7.116	0.576	0.320	1.636	0	1

(Continued)

Table 5 (Continued)

	Successive Reductions of H				Successive Fittings	Successive Inverses			
$(3)_1$	1	0	0	0.423	0.748	1.211	1.807	-0.271	0
$(3)_2$	0	1	0	-0.832	-1.567	1.807	7.635	-0.293	0
$(3)_3$	0	0	1	-0.661	0.218	-0.271	-0.293	0.241	0
$(3)_4$	0	0	0	5.303	0.027	-0.423	0.831	0.661	1
$(4)_1$	1	0	0	0	0.746	1.245	1.741	-0.324	-0.080
$(4)_2$	0	1	0	0	-1.563	1.741	7.765	-0.188	0.157
$(4)_3$	0	0	1	0	0.222	0.324	0.190	0.324	0.125
$(4)_4$	0	0	0	1	0.005	0.080	0.157	0.126	0.189

$$g_1 = (z'_1 \cdot x) = 2.0753$$

$$g_2 = (z'_2 \cdot x) = -0.5999$$

$$g_3 = (z'_3 \cdot x) = 2.5121$$

$$g_4 = (z'_4 \cdot x) = -0.2596$$

Some of the results are given below:

$$b_1 = 0.7460$$

$$b_2 = -1.563$$

$$b_3 = 0.2215$$

$$b_4 = 0.0050$$

In the orthogonal case we have

$$b^*_1 = b_1^{(1)} = 1.2578$$

$$b^*_2 = b_2^{(2)} = -1.301$$

$$b^*_3 = b_3^{(3)} = 0.2182$$

$$b^*_4 = b_4^{(4)} = 0.00502$$

$$c^*_{11} = c_{11}^{(1)} = 0.60606$$

$$c^*_{22} = c_{22}^{(2)} = 7.278$$

$$c^*_{33} = c_{33}^{(3)} = 0.24076$$

$$c^*_{44} = c_{44}^{(4)} = 0.18858$$

Also, we have

$$b^*_o = \bar{x} = 1.9968$$

$$b_o = \bar{x} - \sum_{i=1}^{4} b_i \bar{z}_i = 3.3098$$

The sum of squares assignable to error is SSE = 1.190.

We wish to test whether $\beta_i = 0$ at the 50% level. First, we perform t-tests on the individual b_i. The t ratios are $b_i \sqrt{\frac{20}{(SSE)\ c_{ii}}}$, $i = 1,\ldots, 4$.

For

$$b_0,\ t = 19.3$$

$$b_1,\ t = 2.74$$

$$b_2,\ t = -2.30$$

$$b_3,\ t = 1.59$$

$$b_4,\ t = 0.0474$$

At the 5% level for 20 degrees of freedom t = 2.086. Thus β_0, β_1, and β_2 test to be significantly different from zero, while β_3 and β_4 do not. It remains in doubt whether the hypothesis that both β_3 and β_4 are zero is acceptable, toward this end we shall test whether $\beta_{r+1} = \ldots = \beta_4 = 0$ at the 5% level.

The F-statistic is $\dfrac{\sum_{i=r+1}^{4} b^*_i\ g^*_i\ /(4-r)}{SSE/20}$. We may summarize the tests in the following table.

For Testing H_o	The F Statistic is	While the Critical Value is	Indicating
$\beta_0 = \ldots = \beta_4 = 0$	345.1	$F_{5,20} \doteq 2.7109$	Reject H_o
$\beta_1 = \ldots = \beta_4 = 0$	12.77	$F_{4,20} = 2.8661$	Reject H_o
$\beta_2 = \beta_3 = \beta_4 = 0$	2.411	$F_{3,20} = 3.0984$	Accept H_o
$\beta_3 = \beta_4 = 0$	1.662	$F_{2,20} = 3.4925$	Accept H_o

It thus appears that the data do not strongly support the inclusion of the third and fourth vectors in the regression model.

In the third line of the above table the anomalous situation arises in which we appear to have no grounds (on the basis of the F-test) for rejecting the hypothesis that

$$\beta_2 = \beta_3 = \beta_4 = 0$$

However, in a t test β_2 alone tested to be significantly different from zero. This test should be given precedence. The low F value in the other test is explained by the fact that

$$F = \frac{\left[\dfrac{(b_2^{(2)})^2}{c_{22}^{(2)}} + \dfrac{(b_3^{(3)})^2}{c_{33}^{(3)}} + \dfrac{(b_4^{(4)})^2}{c_{44}^{(4)}}\right] \Big/ 3}{SSE/20}$$

which is approximately equal to $\dfrac{\dfrac{b_2^2}{c_{22}} \Big/ 3}{SSE/20}$, since the contributions

$$\frac{(b_3^{(3)})^2}{c_{33}^{(2)}} + \frac{(b_4^{(4)})^2}{c_{44}^{(4)}}$$

were so small, and hence F was less than t^2.

Finally, if we stop with the second fitting without proceeding to the third and fourth, we would obtain a different SSE (the old SSE with $\dfrac{(b_3^{(3)})^2}{c_{33}^{(3)}}$ and $\dfrac{(b_4^{(4)})^2}{c_{44}^{(4)}}$ "pooled" into it) and a different set of tests. In fact it is easy to see from the bracketed solutions in Tableau Number 2 that $b_1^{(2)}$ and $b_2^{(2)}$ individually and together test out to be significantly different from zero.

There are questions in regression analysis which the above procedure leaves unanswered. One of these is how, out of p predictor variables, we can select a set of k which are "better" than any other set of k. Of course it could be done by considering the regression models for all combinations of k of the p variables and calculating as above the reduction in the sum of square which would be effected for each such regression model; but this is a laborious and time consuming process. An efficient procedure for doing this has not yet been developed.

APPENDIX

THE NORMAL AND RELATED DISTRIBUTIONS

Appended here is a brief development of those distributions which are needed in making the tests discussed in the body of this paper. These distributions are all concerned with statistics which are based on samples from a normal distribution; that is, based on independent random variables each of which is distributed by the same normal distribution law.

A random variable X is said to be normally distributed in case for each real x,

$$F(x) = P\{X \leq x\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{x'-m}{\sigma}\right)^2\right] dx'$$

where m and σ are parameters, $-\infty < m < \infty$, $\sigma > 0$. The mean of such a variable is m and its variance σ^2. For brevity we say X is $N(m, \sigma^2)$. In case X is $N(0,1)$ we say that X is standard normal.

It is well-known (see Cramér (2)) that a distribution is uniquely determined by its characteristic function, which is defined as the Fourier integral, Ee^{itX}. For instance, the characteristic function of the normal variable is

$$Ee^{itX} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[itx - \frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx =$$

$$= \exp\left\{mit - \frac{\sigma^2 t^2}{2}\right\}$$

Another important class of distributions is the class of gamma distributions. A random variable is said to have a gamma distribution in case

$$F(x) = P\{X \leq x\} = 0, \quad x \leq 0$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^x \tilde{x}^{\alpha-1} e^{-\frac{\tilde{x}}{\beta}} d\tilde{x}, \; x > 0$$

where $\alpha > 0$ and $\beta > 0$ are parameters. The characteristic function of a gamma variable is

$$Ee^{itX} = (1 - \beta it)^{-\alpha}$$

We shall now derive the so-called chi-square distribution; that is, the distribution of a sum of squares of independent $N(0,1)$ variables.

First, let X be a $N(0,1)$ variable, and consider $Y = X^2$. The distribution function of Y is $F(y) = P\{Y \leq y\} = P\{X^2 \leq y\}$. This is clearly zero for $y \leq 0$. For $y > 0$

$$F(y) = P\{|X| \leq y^{\frac{1}{2}}\} \frac{1}{\sqrt{2\pi}} \int_{-y^{\frac{1}{2}}}^{y^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2}\right) dx =$$

$$\frac{2}{\sqrt{2\pi}} \int_0^{y^{\frac{1}{2}}} \exp\left(-\frac{x^2}{2}\right) dx$$

Make the transformation $y' = x^2$ to obtain

$$F(y) = \frac{1}{(\pi 2)^{\frac{1}{2}}} \int_0^y y'^{-\frac{1}{2}} e^{-\frac{y'}{2}} dy'$$

Thus $Y = X^2$ is clearly a gamma variable with $\alpha = \frac{1}{2}$, $\beta = 2$. Also such a variable Y has a characteristic function $(1-2it)^{-\frac{1}{2}}$.

Now let $X_1, \ldots, X_k$ be $N(0,1)$ and independent. Clearly $Y_i = X_i^2$ $(i = 1,2,\ldots,k)$ are all distributed as the square of a $N(0,1)$ variable X. They are also independent. Now consider

$$Y = Y_1 + \dots + Y_k = X_1^2 + \dots + X_k^2$$

It can be easily shown that the characteristic function of a sum of independent random variables is the product of their separate characteristic functions. Hence the characteristic function of Y is

$$(1 - 2it)^{-\frac{k}{2}}$$

However, since this is the characteristic function of a gamma variable with $\alpha = \frac{k}{2}$ and $\beta = 2$, it follows that the sum of squares of k independent $N(0,1)$ variables has a gamma distribution with $\alpha = \frac{k}{2}$, $\beta = 2$. Such a distribution is called a chi-square distribution with k degrees of freedom.

Now suppose we have k + 1 independent $N(0,1)$ variables $Y, Y_1, \dots, Y_k$. Let $Z = \sum_{i=1}^{k} Y_i^2$. The random variable Z is clearly a chi-square variable with k degrees of freedom. We define the random variable

$$T = \frac{k^{\frac{1}{2}} Y}{Z^{\frac{1}{2}}} = \frac{Y}{\sqrt{\sum_{i=1}^{k} \frac{Y_i^2}{k}}}$$

In the body of this paper, we saw that such a random variable is important in testing hypotheses on certain parameters whose estimates are normally distributed with unknown variances. The distribution of this statistic, the so-called t-statistic, is derived as follows:

$$F(t) = P\left\{T \leq t\right\} = P\left\{\frac{k^{\frac{1}{2}} Y}{Z^{\frac{1}{2}}} \leq t\right\}$$

Since Y and Z are independent, their joint density function is the following:

$$\varphi\ (y,z) = \frac{1}{\sqrt{2\pi}}\ e^{-y^2}\ \frac{1}{2^{k/2}\ \Gamma\left(\frac{k}{2}\right)}\ z^{\frac{k}{2}-1}\ e^{-\frac{z}{2}}$$

$$= \left(\frac{1}{2^{k+1}}\right)^{\frac{1}{2}}\ \frac{1}{\Gamma\left(\frac{k}{2}\right)}\ z^{\frac{k}{2}-1}\ \exp\left(-y^2 - \frac{z}{2}\right),\quad z > 0$$

$$\varphi\ (y,z) = 0,\quad z \leq 0$$

Hence

$$F(t) = P\left\{z > 0,\ Y \leq \left(\frac{Z}{k}\right)^{\frac{1}{2}}\ t\right\}$$

$$= \left(\frac{1}{2^{k+1}}\right)^{\frac{1}{2}}\ \frac{1}{\Gamma\left(\frac{k}{2}\right)}\ \int_0^{\infty}\left[\int_0^{\left(\frac{z}{k}\right)^{\frac{1}{2}}t} z^{\frac{k}{2}-1}\ e^{-(y^2+\frac{z}{2})}\ dy\right] dz$$

By making a change of variable, $t' = \left(\frac{k}{z}\right)^{\frac{1}{2}} y$, interchanging the order of integration and integrating out the z, this reduces to

$$F(t) = \frac{1}{\sqrt{k\pi}}\ \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\ \int_{-\infty}^{t}\left(1 + \frac{t'^2}{k}\right)^{-\frac{k+1}{2}}\ dt'$$

This is the t-distribution with k degrees of freedom.

Another distribution, used in the body of this paper to test hypotheses on regression weights, is the F-distribution. It is essentially the distribution of the ratio of two independent chi-square variables. This distribution is derived as follows:

Let $X_1, \ldots, X_m$, $Y_1 \ldots Y_n$ be $m + n$ independent $N(0,1)$ variables. Then $X = \sum_{i=1}^{m} X_i^2$ and $Y = \sum_{j=1}^{n} Y_j^2$ are independent chi-square variables with m and n degrees of freedom respectively. Thus, the joint density function of X, Y is

$$\varphi(x,y) = \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{x+y}{2}}, \quad x > 0,\ y > 0$$

$$\varphi(x,y) = 0 \quad \text{otherwise.}$$

Then the distribution of $U = \frac{X}{Y}$ is

$$F(u) = P\left\{0 < X,\ 0 < Y,\ X \leq uY\right\}$$

$$= \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \int_0^\infty \left[\int_0^{uy} x^{\frac{m}{2}-1} y^{\frac{n}{2}-1} e^{-\frac{(x+y)}{2}} dx \right] dy$$

or making the substitution $t = \frac{x}{y}$ and interchanging the order of integration,

$$F(u) = \frac{1}{2^{\frac{m+n}{2}} \Gamma(\frac{m}{2}) \Gamma(\frac{n}{2})} \int_0^u \int_0^\infty t^{\frac{m}{2}-1} y^{\frac{m+n}{2}-1} e^{-\frac{(t+1)y}{2}} dy\, dt$$

However, since

$$\int_0^\infty y^{\frac{m+n}{2}-1} e^{-\frac{(t+1)y}{2}} dy = 2^{\frac{m+n}{2}} \Gamma(\tfrac{m+n}{2}) \frac{1}{(t+1)^{\frac{m+n}{2}}},$$

then

$$F(u) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\,\Gamma\left(\frac{n}{2}\right)} \int_0^u \frac{t^{\frac{m}{2}-1}}{(t+1)^{\frac{m+n}{2}}}\,dt$$

Then the variable

$$\mathcal{F} = \frac{X/m}{Y/n} = \frac{n}{m}\,U$$

has the distribution

$$P\{\mathcal{F} \leq F\} = G(F) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{n}\right)\,\Gamma\left(\frac{n}{2}\right)} \int_0^{\frac{m}{n}F} \frac{t^{\frac{m}{2}-1}}{(t+1)^{\frac{m+n}{2}}}\,dt$$

whereupon, making the change of variable $v = \frac{n}{m}\,t$

$$G(F) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \int_0^F \frac{v^{\frac{m}{2}-1}}{\left(\frac{m}{n}\,v+1\right)^{\frac{m+n}{2}}}\,dv \;.$$

BIBLIOGRAPHY

1. Anderson, R. L. and T. A. Bancroft, Statistical Theory in Research. New York, McGraw-Hill Book Company, 1952.

2. Cramer, H., Mathematical Methods of Statistics. Princeton, Princeton University Press, 1957.

3. Fraser, D. A. S., Statistics, an Introduction. New York, John Wiley and Sons, Inc., 1958.

4. Mood, A. M., Introduction to the Theory of Statistics, New York, McGraw Hill Book Company, 1950.

5. Murdoch, D. C., Linear Algebra for Undergraduates New York, John Wiley and Sons, Inc., 1957.

6. Scheffe, H., The Analysis of Variance. New York, John Wiley and Sons, Inc., 1959.

7. Wald, A., "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations is Large." Transactions of the American Mathematical Society, Vol. 54 (1943), pp. 426-482.

8. Williams, E. J., Regression Analysis. New York, John Wiley and Sons, Inc., 1959.

9. Woltz, W. G. Reid, W. A., and Colwell, W. E., "Sugar and Nicotine in Cured Bright Tobacco as Related to Mineral Element Composition." Soil Science Society of America Proceedings, Vol. 13 (1948), pp. 385-387.